MESOSCALE SYSTEMS, FINITE SIZE EFFECTS, AND BALANCED NEURAL NETWORKS

by

J. B. Dunworth

B.S. Mathematics, Physics, UCLA, 2006

M.S. Physics, University of California, San Diego, 2008

M.S. Mathematical Biology, The Ohio State University, 2011

Submitted to the Graduate Faculty of

the Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

J. B. Dunworth

It was defended on

July 24th 2018

and approved by

Brent Doiron, Department of Mathematics

G. Bard Ermentrout, Distinguished University Professor, Depts. of Mathematics &

Computational Biology

Jonathan Rubin, Dept. of Mathematics

Anne-Marie Oswald, Dept. of Neuroscience

Dissertation Director: Brent Doiron, Department of Mathematics

MESOSCALE SYSTEMS, FINITE SIZE EFFECTS, AND BALANCED NEURAL NETWORKS

J. B. Dunworth, PhD

University of Pittsburgh, 2019

Cortical populations are typically in an asynchronous state, sporadically interrupted by brief epochs of coordinated population activity. Current cortical models are at a loss to explain this combination of states. At one extreme are network models where recurrent inhibition dynamically stabilizes an asynchronous low activity state. While these networks are widely used they cannot produce the coherent population-wide activity that is reported in a variety of datasets. At the other extreme are models where short term synaptic depression between excitatory neurons can generate the epochs of population-wide activity. However, in these networks inhibition plays only a perfunctory role in network stability, which is at odds with many reports across cortex. In this study we analyze spontaneously active *in vitro* preparations of primary auditory cortex that show dynamics that are emblematic of this mixture of states. To capture this complex population activity we consider models where large excitation is balanced by recurrent inhibition yet we include short term synaptic depression dynamics of the excitatory connections. This model gives very rich nonlinear behavior that mimics the core features of the *in vitro* data, including the possibility of low frequency (2-12 Hz) rhythmic dynamics within population events. Our study extends balanced network models to account for nonlinear, population-wide correlated activity, thereby providing a critical step in a mechanistic theory of realistic cortical activity. We further investigate an extension of this model that l exhibits clearly non-Arrhenius behavior, whereby lower noise systems may exhibit faster escape from a stable state. We show that this behavior is due to the system size dependent vector field, intrinsically linking noise and dynamics.

TABLE OF CONTENTS

PREFACE ix

1.0 INTRODUCTION 1

1.1 Historical mean field models 1

1.2 The Wilson-Cowan model 3

1.3 Neural variability 5

1.3.1 Balanced networks 6

1.4 Synaptic plasticity 9

1.4.1 A simplified biological explanation of plasticity 9

1.4.2 The canonical synapse model 10

1.4.3 Integrating STP into a network 13

1.5 Thesis outline 14

2.0 INTERLEAVING ASYNCHRONOUS AND SYNCHRONOUS ACTIVITY IN BALANCED CORTICAL NETWORKS WITH SHORT TERM SYNAPTIC DEPRESSION 16

2.1 Results 18

2.1.1 Two kinds of network stabilization 18

2.1.2 Oscillatory and non-oscillatory population events 22

2.1.3 Balanced cortical circuit model of population events 23

2.1.4 Population event initiation 31

2.1.5 Population event rhythmicity 34

2.1.6 Suite of population event dynamics 34

2.2 Discussion 37

2.2.1 Synaptic depression and population events in the auditory cortex . . . 40
2.2.2 Balanced networks and population events . . . 42
2.3 Methods . . . 43
2.3.1 Markov model of network . . . 43
2.3.2 Model of a synapse . . . 46
2.3.3 Simulation . . . 46
2.3.4 Bifurcation diagrams . . . 47
2.3.5 Event detection in data . . . 48
3.0 ESCAPE PROBLEMS . . . 50
3.1 Introduction . . . 50
3.2 Mathematical framework . . . 54
3.2.1 Canonical escape example . . . 59
3.3 Neuronal model with synaptic plasticity . . . 66
3.4 Comparison between scaling laws . . . 71
3.5 Stochastic hybrid systems and the pseudopotential . . . 72
3.5.1 Quasi-steady state approximation . . . 74
3.5.2 Escape time estimation . . . 78
3.5.3 Eigenvalue estimation . . . 84
3.6 Conclusion . . . 88
4.0 CONCLUSION . . . 90
4.1 Spiking simulations . . . 90
4.2 Synaptic model . . . 91
4.3 Parameter choices . . . 92
4.4 Escape time estimate . . . 94
REFERENCES . . . 95

LIST OF TABLES

1 Commonly used symbols x
2 Chapter 2 parameters 45
3 Chapter 3 parameters 68

LIST OF FIGURES

1 System size dependent nullclines 9
2 Synaptic plasticity overview 11
3 Two mechanisms of stabilization 21
4 Heterogeneity of dynamics of population events 24
5 Balanced network model with synaptic depression of excitatory connections . 26
6 Birth-death Markov model of network activity 29
7 Noise due to finite network size induces population events 30
8 Reduced network model with fixed EE synaptic strength 33
9 Reduced network model with fixed EI synaptic strength 35
10 Suite of network dynamics 38
11 Detailed bifurcation structure 39
12 Toy model of transitions 53
13 Escape from a double well 61
14 Scaling law comparison 70
15 Reduced model calculation 77
16 Theory vs. simulation comparison 83
17 Synaptic model fit to data 93

PREFACE

I am deeply indebted to many people without whom this thesis would not have been completed. I would like to thank the University of Pittsburgh College of Letters and Sciences, the National Institutes of Health, the National Science Foundation, and the Benter Foundation for their generous support during my course of study. I would also like to thank my thesis committee, Drs. Brent Doiron, G. Bard Ermentrout, Jon Rubin, and Anne-Marie Oswald, who provided guidance throughout this thesis. Special thanks go to the Ermentrout and Doiron labs for many helpful conversations and direction. I have been lucky to have had many helpful officemates and peers to use as a sounding board and a sanity check, and my work would have suffered if not for the feedback and encouragement of Glenn Young, Jeremy Harris, and Jay Pina. Drs. Chengcheng Huang and Hannah Bos provided numerous helpful learning experiences, and my work is clearer and simpler for their input.

My deepest thanks to my family for their constant support and love. Special mention should be given to my wife, Marisa Eisenberg, for not only providing me with support and encouragement, but also for her tireless patience as a sounding board, and for not leaving me during the writing phase, when (I assure you) I was extremely unpleasant to deal with.

Finally, the role of my thesis advisors cannot be overstated. The best ideas in my work have come through conversations with my advisors Brent and Bard, and I am extremely fortunate that they let me work through the bad ones on my own to see where my work could be improved. They have both been friends throughout my studies, and I have learned much from them both, not only about mathematics and computational neuroscience, but about careers in research and the challenges (scientific and not) that accompany scientific pursuit. This work could not have been completed without their guidance, and I can't imagine a better set of mentors from which to learn – my many thanks to them both.

Symbol	Typical Meaning	State Space
r	firing rate	$\mathbb{R}$
n	number of binary neurons in up state	$[0, N] \subset \mathbb{N}$
N	system size, number of neurons	$\mathbb{N}$
K	average number of connections between neurons	$\mathbb{N}$
p	plasticity variable, synaptic efficacy	$[0, 1] \subset \mathbb{R}$
τ	timescale	$\mathbb{R}$
J	absolute connection strength between neurons	$\mathbb{R}$
j	scaled connection strength between neurons	$\mathbb{R}$
I	applied current	$\mathbb{R}$
f	sigmoidal function	$[0, 1] \subset \mathbb{R}$
θ	threshold parameter	$\mathbb{R}$
ρ	probability density	$[0, \infty) \subset \mathbb{R}$
Ψ	vector of probability densities	$\mathbb{R}^m$
$\mathcal{S}$	survival probability	$[0, 1] \subset \mathbb{R}$
$\mathcal{N}$	normalization constant	$\mathbb{R}$
$\mathcal{U}$	Potential function	$\mathbb{R}$
$\hat{\mathcal{U}}$	Pseudo-potential function	$\mathbb{R}$
σ	characteristic amplitude of white noise	$\mathbb{R}$
ω	transition rate (Markov process)	$\mathbb{R}$
λ, μ(Chapter 3)	eigenvalues	$\mathbb{R}$
η, φ	eigenvectors	$\mathbb{R}^m$

Table 1: Commonly used symbols

1.0 INTRODUCTION

1.1 HISTORICAL MEAN FIELD MODELS

Neuroscience is a fundamentally multiscale problem, wherein millions of individual neurons (already complex individually) interact to form a functioning, highly complex whole. This makes it highly challenging to model individual neurons and neuronal interactions, and so nearly all models make strong simplifying assumptions in order to make capturing the larger interaction patterns tractable[1, 2]. Moreover, the complexity of single neuron models varies quite a bit depending on the goals of the modeling and the questions to be addressed.

For example, the earliest qualitative model of a neuron comes from Lapique[3] with the *leaky integrate-and-fire neuron*, which incorporates threshold spiking, decay to a resting membrane potential, and a timescale of integration, features used in almost all models of neurons. There have since been developed a wide range of dynamical extensions of this basic model. Some extensions are developed to model specific types of neurons[4], some for their interesting dynamical character[5, 6], while others have been developed for their ease of computation[7] and usefulness in formalizing the essential characteristics of complex behavior[8].

The gold standard in single neuron modeling is the Hodgkin Huxley neuron [9, 10], work which led to a Nobel prize in physiology and medicine in 1963, one of the few examples of such a Nobel awarded for mathematical biological work. Of particular note is that the development, parameter estimation, and model comparison done using this model allowed Hodgkin and Huxley to predict the existence and number of sodium and potassium gates within the neuron before they were validated experimentally. The Hodgkin-Huxley model is one of the most celebrated examples of the useful role of mathematical modeling in predicting

and testing potential biological mechanisms. It continues to be widely used in practice today, with a wide range of variations and extensions developed for different cell and tissue types[11–14].

However, in spite of the successful development of models for single neurons across complexity scales, modeling larger interaction patterns in the brain remains a daunting challenge. Brains consist of billions of neurons wired together, and even networks composed of the simplest models of neurons can lead to seemingly arbitrarily complicated dynamics, depending on the variables considered. Indeed, the question of understanding behavior of neural networks has been studied since the earliest days of computational biology [15], and an agreed-upon or "correct" approach to modeling aggregate activity remains elusive. Models of network activity have historically relied heavily on the use of field theories, a well-defined mathematical object allowing for rigorous analysis of dynamical character. Early models, some of which pre-date our knowledge of inhibitory interneurons, discuss the stability characteristics of networks of pulse coupled units[16–18], and the field eventually settled on mean field models of recurrently coupled excitatory and inhibitory networks as the minimal necessary model to understand neural activity[19, 20]. The currently accepted standard, or canonical model, is based on the work of Wilson & Cowan[21], where the authors derive a set of field equations for recurrently connected excitatory inhibitory networks.

Field theories are particularly useful for modeling the brain for several reasons. First, many of the best measurements available of large scale neural activity are aggregate measures, which inherently involve some sort of averaging process,such as local field potential or electroencephalogram recordings[22, 23]. As such, there is an inherent dimension reduction necessary for a model to be validated against measurements, which makes it a sensible choice to simply model the aggregate behavior directly. Second, though neural data can be arbitrarily complex (depending on the scale examined), there often appears to be a lower dimensional description that provides the same functional responses [2]. Given that the aggregate activity seems as though it is actually on a low dimensional manifold, a description of the lower dimensional behavior is attractive. Finally, brains are formed by vast numbers of (relatively tiny) neurons (e.g. humans have $\sim 10^{11}$ neurons, densely connected). This suggests that models of even moderately sized regions (e.g. a few centimeters) will contains

thousands and thousands of neurons, and the different operating scales of individual units vs. observable activity suggests that continuum style models would be an appropriate modeling choice. These methods have been extremely successful in the study of propagation of wave-like activity[24] and spatiotemporal pattern formation[25, 26].

The work of Wilson & Cowan has been highly influential[27], and has led to the general advancement of computational neuroscience and the role of mathematics and dynamical systems, especially in the study of neural systems. Nonetheless, there remain a host of open questions surrounding the optimal approach and level of detail needed when modeling neural systems, even discounting the incredibly complicated questions of how computations are performed for more complex models. Since the development of the Wilson-Cowan model, there have been many extensions of the work of studying neural field models, and the current standard techniques involve statistical mechanics[28–30], theories of correlations [31, 32], the study of complicated oscillations[33], and chaos[34]. More recently, extensions to standard mean field models have been developed, which attempt to robustly account for variability in systems[35, 36].

Here, I examine a specific extension of existing neuronal mean field models, which allows for internally generated variability (i.e. we are able to quell the need for uncorrelated white noise in simulations to ensure asynchrony[29]), a key feature observed in a wide range of neuronal activity recordings [37]. This extension was developed by van Vreeswijk & Sompolinsky [38], and is termed a *balanced network*. Below we explore some of the key features of the Wilson-Cowan model, and the balanced network extension that we will study in the remainder of this work.

1.2 THE WILSON-COWAN MODEL

We present here a brief introduction to the Wilson-Cowan rate model. The argument we present here is not the original argument provided by Wilson and Cowan, but rather is based on an argument first presented by Ginzburg & Sompolinsky [31], based on the techniques of statistical mechanics and Glauber dynamics[39]. We believe this derivation highlights a

salient point we would like to draw attention to, though the Wilson-Cowan theory is so well established that there are a number of alternative derivations[2]. Let our network consist of N binary neurons, with the state of the i^{th} neuron at time t denoted by $s_i(t)$. As this is a model of binary neurons, $s_i \in \{0, 1\}$. The state s_i is determined as a threshold of a linear combination of its inputs, given by

$$s_i(t) = \Theta\left(\sum_{j=1}^{N} J_{ij} s_j(t) + \mu^b - \theta\right) \tag{1.1}$$

where μ^b is a constant input to all neurons (analogous to an input current), θ is the threshold, J_{ij} is the strength of the connection from neuron j to neuron i, and $\Theta(\cdot)$ is the Heaviside function. It is assumed that $J_{ii} = 0 \, \forall i = 1, \ldots, N$. The neurons are partitioned into populations of excitatory neurons and inhibitory neurons, and we assume that the average rate of update for neurons in population α is given by the time constant $1/\tau_\alpha$. Note we allow for heterogeneity in the inputs and thresholds, meaning we have μ^b_α and θ_α, $\alpha \in \{E, I\}$. For simplicity, we may take J_{ij} to be one of four constants, representing the strength of connection between neuron *types*, rather than specific neurons. We denote strength of a connection from a neuron in population β to a neuron in population α as $J_{\alpha\beta}$, where $\alpha, \beta \in \{E, I\}$. Allowing heterogeneity among the J_{ij} by, for example, taking them as drawn from a Gaussian distribution, as in Wilson & Cowan [21], does not appreciably change the derivation.

We further assume a network connectivity where any two neurons are statistically indistinguishable (as would be the case in all-to-all networks, K-regular networks, and Erdös-Renyi networks), and that the activity is roughly asynchronous. Let an arbitrary neuron receive, on average, K connections from both the E and I populations. It can be shown [28, 31, 39] that the average activity of population α, $m_\alpha(t)$, will be given by

$$\tau_\alpha \frac{dm_\alpha}{dt}(t) = -m_\alpha(t) + \text{erfc}\left(\frac{-\mu_\alpha}{\sqrt{V_\alpha}}\right) \tag{1.2}$$

$$\mu_\alpha = J_{\alpha E} K m_E - J_{\alpha I} K m_I + \mu^b_\alpha - \theta_\alpha \tag{1.3}$$

$$V_\alpha = J^2_{\alpha E} K m_E + J^2_{\alpha I} K m_I + V^b_\alpha \tag{1.4}$$

where $\text{erfc}(x)$ is the complementary error function, μ_α is the mean input into population α, V_α is the mean variance of inputs into population α, and V^b_α is the background variance

intrinsic to the population α.

To recover the classical formulation of the Wilson-Cowan equations, we must make a few further simplifications. First, we let $K \to \infty$. To prevent the mean input into a neuron from diverging, we assume the connection strengths may be written as $J_{\alpha\beta} = j_{\alpha\beta}/K$, where $j_{\alpha\beta} \sim \mathcal{O}(1)$. Under this constraint, the variance $V_\alpha \to V_\alpha^b$ in the $K \to \infty$ limit, a constant. Finally, we may define a sigmoidal function $f(x)$ that absorbs the background variance and a negative sign to arrive at

$$\tau_\alpha \frac{dm_\alpha}{dt} = -m_\alpha + f\left(j_{\alpha E} m_E - j_{\alpha I} m_I + \mu_\alpha^b - \theta_\alpha\right), \quad \alpha \in \{E, I\} \tag{1.5}$$

If we associate the firing rate of a population with its mean activity, we have constructed a pair of coupled rate equations with sigmoidal firing rate functions for a recurrently coupled exictatory-inhibitory network. The techniques presented here are extendable to more complicated models of single neurons, but we feel that this model is sufficient to highlight the essential characteristics of the Wilson-Cowan style rate models.

1.3 NEURAL VARIABILITY

As we saw in section 1.1, one expects to see some variability in the firing rates of a model of neural activity. In our construction of the Wilson-Cowan model in section 1.2, we showed that the variance of the neuronal inputs limits to the background variance in the limit of a network of many neurons. To get variable firing rates from such a system, one suspects we would need quite a large background variance. Softky & Koch [40] formalized this argument as follows.

Actual physiological neurons integrate inputs, which takes time. Neurons also have a threshold of some kind that must be crossed to initiate an action potential. Let us take as a model neuron a perfect integrator, and we suppose that the connection strengths are scaled as they are in section 1.2, with $J \sim 1/K$. Then to initiate an action potential, a neuron needs to receive K excitatory inputs. We know that there is, at a minimum, some kind of background noise in the system, so we may assume that our perfect integrator is receiving

excitatory inputs from a Poisson process with rate r. Then the time between two spikes of our perfect integrator is the *interspike interval*, a random variable that we can quantify. Given a Poisson input, the expected value of the interspike interval will be K/r, with a standard deviation of $\sqrt{K}/r$. The standard method of quantifying variability in spike trains is the *coefficient of variation* of the distribution of interspike intervals, given as the ratio of the standard deviation to the mean. For this toy system, the coefficient of variation will be $1/\sqrt{K}$. Then as $K \to \infty$, the coefficient of variation approaches zero, giving a highly regular spike train.

To rectify this incongruity with observed data, Shadlen & Newsome [41] proposed a solution in the form of excitatory/inhibitory balance. If an integrator receives both excitatory and inhibitory inputs of (approximately) the same size, the voltages should look more like a random walk, meaning that the times at which the integrator reaches the spike initiation threshold should be more variable than in Softky & Koch [40].

There is ample evidence of excitatory/inhibitory balance from experiment[1]. Some form of balance between excitation and inhibition has been seen in slice[42], *in vitro* [43], *in vivo*[44–47], and in culture[48]. The culture validation is particularly exciting as it validates a modeling choice that we will see. Additionally, dysregulation of excitatory/inhibitory balance has been implicated in some of disease states[49–53].

1.3.1 Balanced networks

We are now ready to extend the Wilson-Cowan rate model framework to allow for internally generated variability. Beginning from the same system as in section 1.2, we can follow the same calculations through Eq. (1.1) to (1.4). At the level of Eq. (1.3) and (1.4), though, we make a different ansatz. Motivated by the desire to retain variance other than the background variance V_b, we choose to scale the connection strengths as $J_{\alpha\beta} = j_{\alpha\beta}/\sqrt{K}$, where $j_{\alpha\beta} \sim \mathcal{O}(1)$. This results in $V_\alpha \sim \mathcal{O}(1)$, giving internally generated variability through the dynamic recurrent interactions between and among the two neuronal populations. A side effect of this, however, is that $\mu_\alpha \sim \mathcal{O}\left(\sqrt{K}\right)$. Arbitrarily large inputs will lead to saturation

[1] Though we note there is much more evidence now than when Shadlen & Newsome proposed the idea.

of the neuronal activity.

To prevent this, we make the following simplifications. First, assume the threshold θ_α is negligible. Next, assume that the input μ_α^b is large, so that it may also be written as $\mu_\alpha^b = I_\alpha\sqrt{K}$. If we are thinking of μ_α^b as input from another brain area, this is equivalent to the statement that feedforward connections can be strong. With these simplifications, Eq. (1.3) becomes

$$\mu_\alpha = \sqrt{K}\left(j_{\alpha E}m_{_E} - j_{\alpha I}m_{_I} + I_\alpha\right) \tag{1.6}$$

To ensure that $\mu_{_E}, \mu_{_I} \sim \mathcal{O}(1)$, it is enough require that

$$j_{\alpha E}m_{_E} - j_{\alpha I}m_{_I} + I_\alpha \sim \left(\frac{1}{\sqrt{K}}\right). \tag{1.7}$$

In the $K \to \infty$ limit, this gives rise to the balance conditions

$$j_{EE}m_{_E} - j_{EI}m_{_I} + I_{_E} = 0 \tag{1.8}$$

$$j_{IE}m_{_E} - j_{II}m_{_I} + I_{_I} = 0 \tag{1.9}$$

The balance conditions are generically solvable, and they provide a set of inequalities we require if we expect the activities for our model to be well-defined. Note that in our desire to keep the total inputs variable (i.e. $V_\alpha \sim \mathcal{O}(1)$), we have derived, without explicit intent, the condition outlined by Shadlen & Newsome [41].

Implicit in the derivation we provided here is the assumption that the network is sparse, meaning that even though both K and N diverge to infinity, $K \ll N$ so that the ratio $K/N \to 0$. This assumption is tantamount to the statement than any two neurons in the network are at best weakly correlated. This is effectively an asynchrony condition. Renart et al. [54] extended the argument in van Vreeswijk & Sompolinsky [55] to allow balanced states in dense networks, where the ration K/N is a nonzero constant. That argument was later simplified and extended to allow for spatial heterogeneity[56]. A result of this extension to dense networks provides that balanced networks actively decorrelate neuronal activity through a dynamic tension in the afferent currents onto a neuron. When a large excitatory current is incoming, the network, on average, provides a compensatory inhibitory current, effectively cancelling the excitatory input. While this places the net input into a

neuron near zero, it still allows fluctuations in the network activity to cause spikes, because the inputs of each type (excitatory/inhibitory) are large. This places the neuron in the *fluctuation driven regime* (as opposed to the oscillatory regime), and suggests the tools of studying noisy systems may be appropriate for analysis of systems of this type.

Finally, we should remark on the dynamics of balanced networks. First, we note that Eq. (1.8) and (1.9) will generically have only a single solution, meaning that the dynamical system of the activity variables has *only a single fixed point* in the $K \to \infty$ limit. Moreover, the dynamics will be governed by Eq. (1.2) to (1.4). Next, we observe that by construction, the input to the sigmoidal function will be

$$\frac{\mu_\alpha}{V_\alpha} \sim \mathcal{O}\left(\sqrt{K}\right) \tag{1.10}$$

which implies that the eigenvalues of the Jacobian at a fixed point will scale with $\sqrt{K}$. Thus in the large K limit, the dynamical system responds to perturbations on a timescale that scales with K: the larger the K, the faster the response. Colloquially put, a stable fixed point becomes *extremely* stable as K grows.

For much of the work that follows, we will adopt a simplified form of a balanced network, where we retain the sigmoidal nature of the transfer function, as well as the $\mathcal{O}\left(\sqrt{K}\right)$ sized inputs. The simplest form of this simplified model is given as

$$\begin{aligned} \tau_{_E}\dot{m}_{_E} &= -m_{_E} + f\left(\sqrt{K}\left(j_{_{EE}}m_{_E} - j_{_{EI}}m_{_I} + I_{_E}\right)\right) \\ \tau_{_I}\dot{m}_{_I} &= -m_{_I} + f\left(\sqrt{K}\left(j_{_{IE}}m_{_E} - j_{_{II}}m_{_I} + I_{_I}\right)\right) \end{aligned} \tag{1.11}$$

where we have used a dot to denote differentiation with respect to time. We note that as the governing equations depend explicitly on K, for any *fixed* K, it is possible to have more diverse dynamical possibilities. In particular, we are able to choose parameters such that Eq. (1.11) exhibit bistability (Figure 1A.) As K grows, the nullclines approach the pair of lines defined by the balance condition (Figure 1B,C). So while it is technically true that a finite sized balanced network can exhibit nonlinear behavior, this is a finely tuned case, and we seek a more robust method to extend the theory of balance. To that end, we follow the work of Mongillo et al. [57], where the authors added robust multistability to balanced networks through the use of short-term synaptic plasticity.

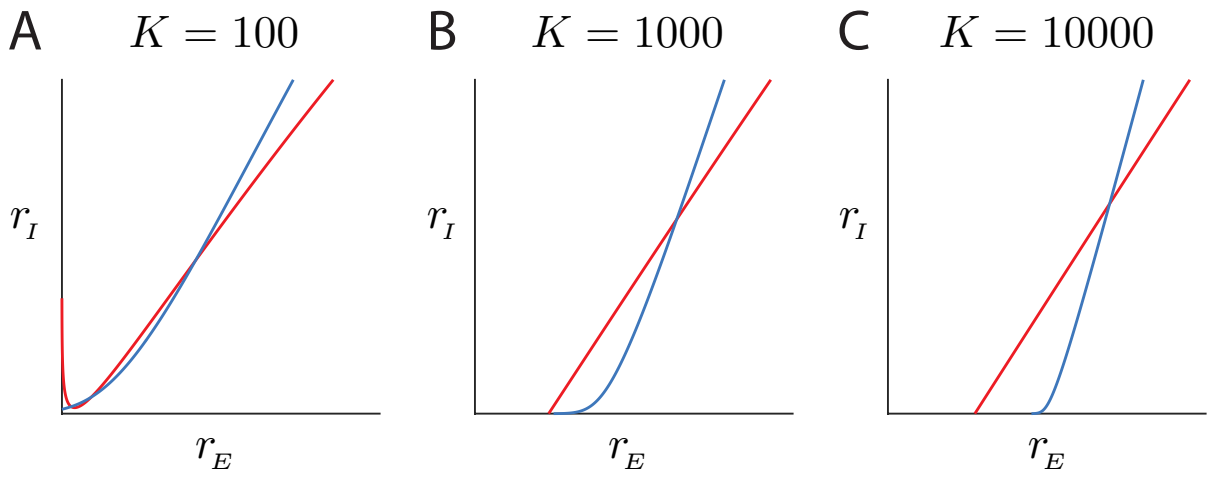

Figure 1: Nullclines for a finite sized balanced network at 3 different system sizes.

1.4 SYNAPTIC PLASTICITY

What follows is a grossly over-simplified explanation of short-term plasticity (STP) in neuronal populations. STP depends on a number of subtle factors, having different characteristics depending on the cell types involved, which regions of cortex are being studies, and the specific details of the experimental preparation. We present here the features most relevant to the modeling work we do in Chapters 2 and 3. For a more comprehensive view, see the reviews [58, 59].

1.4.1 A simplified biological explanation of plasticity

While a detailed biological description of short term plasticity is outside the scope of this work, it is beneficial to give a brief overview. At a very basic level, we may understand the procedure of signal propagation across a synapse as follows. When a presynaptic neuron fires an action potential, synaptic vesicles release neurotransmitter into the area between the synaptic bouton and the post-synaptic neuron. The neurotransmitter binds to receptors on the post-synaptic membrane, causing a temporary change in conductance, and a measurable

post-synaptic potential (PSP). After release, there is a refractory period of several hundred milliseconds during which the synaptic contact does not have a vesicle docked. Under repetitive stimulation, the size of the PSPs changes, either decreasing (termed *depression*) or increasing (termed *facilitation*). These changes depend systematically on stimulus intensity and frequency. The type of STP depends on the types of pre- and postsynaptic cells. Direct measurements, through simultaneous triple whole cell recordings, have shown that a single pre-synaptic neuron can cause depression in one postsynaptic target and facilitation in another[60].[2]

The cellular mechanisms by which STP takes place are better understood for depression than for facilitation, but workable theories for both phenomena exist. The characteristics of depressing synapses are consistent with a simple depletion model, where the readily releasable pool of vesicles decreases with every incoming action potential. The most prevalent theory for facilitation is the residual calcium hypothesis[61], which posits that the probability of vesicle release increases in a nonlinear way with the accumulation of calcium [Ca2+].

1.4.2 The canonical synapse model

There are a variety of models of synapses, ranging from the highly mechanistic to the functionally qualitative, encompassing stochastic effects at the microscale level[62, 63], and network effects arising from large scale interactions of many agents [64–66]. We present here one of the most adaptable and well-regarded models of a synapse[64, 67]. While this model is partly phenomenological, it is known to reproduce well the measured synaptic responses between pyramidal neurons.

We consider a synapse as having a finite amount of resources (e.g. total amount of neurotransmitter), which can be partitioned into three states. The resources can be *active*, or effective (e.g. if a vesicle is docked), *inactive*, or *recovered*. The fraction of active, inactive, and recovered resources is denoted by E, I, and R, respectively. Because the total amount

[2] As an applied mathematician who depends on the data collection of others, we should remark here that direct measurement of STP is a herculean task, one which requires simultaneous patch camp techniques onto at least two neurons. Moreover, one needs a direct connection between two neurons to perform this experiment. The method of finding such a direct connection can best be described as "guess and check" (or, we suppose, "patch and pray"). The skill and patience required for simultaneous triple patching onto divergent connections can likely not be overstated.

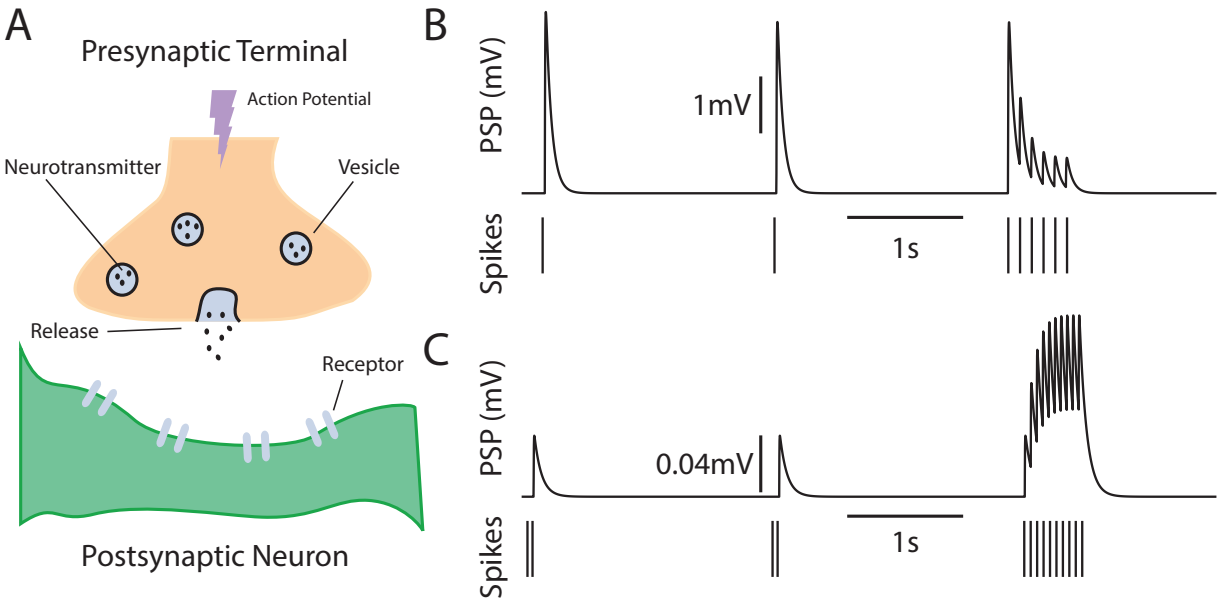

Figure 2: **A:** A cartoon of a synapse. **B:** A presynaptic spike train induces depression in the postsynaptic potential if the spikes are temporally close. Simulation. **C:** A presynaptic spike train induces facilitation in the postsynaptic potential if the spikes are temporally close. Simulation.

of resources is fixed, we have the constraint $E + I + R = 1$.

Let t' be the time of a pre-synaptic action potential, and let U_{SE} (the utilization of synaptic efficacy) be the fraction of synaptic resources moved into the active classification at t'. The timescale with which resources move from active to inactive is a few milliseconds, denoted by τ_{in}, and the timescale with which they recover is one the order of hundreds of milliseconds, up to a second, denoted by τ_{rec}. The kinetic equations for a depressing synapse follow essentially mass action compartmental model, and are found as

$$\begin{aligned} \frac{dR}{dt} &= \frac{I}{\tau_{rec}} - U_{SE} \cdot R(t' - 0) \cdot \delta(t - t') \\ \frac{dE}{dt} &= -\frac{E}{\tau_{in}} + U_{SE} \cdot R(t' - 0) \cdot \delta(t - t') \\ I &= 1 - R - E \end{aligned} \tag{1.12}$$

There is also the parameter A_{SE}, the absolute synaptic efficacy, defined as the maximal possible response if all of the resources are activated by a presynaptic action potential. We take the postsynaptic current, I_p to be proportional to the fraction of resources in the active state ($I_p(t) = A_{SE}E(t)$). While simple, this model effectively captures short term depression (Figure 2B). In particular, we note that it reproduces the observed phenomena that for low presynaptic firing rates depression is almost non-existent (the first two spikes in Figure 2B), but for high firing rates depression is observed (the last five spikes in Figure 2B).

To account for facilitation, only a small modification of this model is needed. Rather than assuming that a fixed amount of resources are released with every action potential, we assume that the effective value of U_{SE} is increased with every action potential. Let U' be the amount of synaptic resources available to be released (the running effective value of U_{SE}, essentially), and let τ_f the time constant of facilitation. Then a model that captures both

facilitation and depression can be enumerated as

$$\begin{aligned}
\frac{dR}{dt} &= \frac{I}{\tau_{rec}} - U' \cdot R(t'-0) \cdot \delta(t-t') \\
\frac{dE}{dt} &= -\frac{E}{\tau_{in}} + U' \cdot R(t'-0) \cdot \delta(t-t') \\
\frac{dU'}{dt} &= -\frac{U'}{\tau_f} + U_{SE} \cdot (1-U') \cdot \cdot \delta(t-t') \\
I &= 1 - R - E
\end{aligned} \tag{1.13}$$

See Figure 2C for an example simulation of facilitation.

1.4.3 Integrating STP into a network

For our purposes, the most important result from this work is a roadmap on how to include synaptic effects in a recurrent network using the mean value of the synaptic resources in the recovered category, $\langle R \rangle$. Following the results of Softky & Koch [40], it is reasonable to model the incoming spike train to Eq. (1.12) and (1.13) as a Poisson process, say with rate $r(t)$. Taking τ_{in} to be much faster than τ_{rec}, we can average over possible realizations of this spike train, and find the following equations:

$$\begin{aligned}
\frac{d\langle R \rangle}{dt} &= \frac{1-\langle R \rangle}{\tau_{rec}} - \langle U' \rangle \cdot \langle R \rangle \cdot r(t) \\
\frac{d\langle E \rangle}{dt} &= -\frac{\langle E \rangle}{\tau_{in}} + \langle U' \rangle \cdot \langle R \rangle \cdot r(t) \\
\frac{d\langle U^- \rangle}{dt} &= -\frac{\langle U^- \rangle}{\tau_f} + U_{SE} \cdot (1 - \langle U^- \rangle) \cdot r(t) \\
\langle U' \rangle &= \langle U^- \rangle (1 - U_{SE}) + U_{SE}
\end{aligned} \tag{1.14}$$

where $\langle U^- \rangle$ is the average value of U' immediately preceding an action potential. This formulation gives us a sense of the typical dynamics of synapse.

To integrate this formulation into a recurrent network of many neurons, we need to compute the postsynaptic current. Recalling that $I_p(t) = A_{SE}E(t)$, we will have the average postsynaptic current as $\langle I_p \rangle = A_{SE}\langle E \rangle$. As a first pass at simplification, we may suppose that τ_{in} is fast relative to the other timescales in the system. Even if τ_{in} is a few milliseconds,

this may be reasonable if we are considering the effective time constant of the excitatory population to be, say, 10ms. In this case, when we are only interested in the behavior of system at timescales slower that τ_{in}, we find the postsynaptic current can be simplified to

$$\langle I_p(t)\rangle = \tau_{in} A_{SE} U' \cdot \langle R\rangle \cdot r(t), \tag{1.15}$$

effectively a product of the presynaptic firing rate and the available proportion of synaptic resources.

As an explicit example, suppose we take a system of the form of Eq. (1.5), and we allow a single synapse type to be plastic, say the excitatory to inhibitory ($E \to I$) connection. Denote the synaptic efficacy as p, a number residing in the unit interval $[0, 1]$. Then a proper accounting of STP in the mean field model is incorporated as

$$\begin{aligned} \tau_{_E}\dot{m}_{_E} &= -m_{_E} + f\left(\sqrt{K}\left(j_{EE}m_{_E} - j_{EI}m_{_I} + I_{_E}\right)\right) \\ \tau_{_I}\dot{m}_{_I} &= -m_{_I} + f\left(\sqrt{K}\left(j_{IE}\cdot p\cdot m_{_E} - j_{II}m_{_I} + I_{_I}\right)\right) \end{aligned} \tag{1.16}$$

where the governing equation for p follows from Eq. (1.14), giving

$$\begin{aligned} \dot{p} &= \frac{1-p}{\tau_{rec}} - \left(\langle U^-\rangle\left(1-U_{SE}\right) + U_{SE}\right)\cdot p\cdot m_{_E} \\ \frac{d\langle U^-\rangle}{dt} &= -\frac{\langle U^-\rangle}{\tau_f} + U_{SE}\left(1-\langle U^-\rangle\right)m_{_E} \end{aligned} \tag{1.17}$$

In the work that follows, we will make a few simplifications motivated by these results, providing a qualitatively similar functional form of a synapse.

1.5 THESIS OUTLINE

In this thesis, we present the work of two studies, both involving the study of balanced neuronal networks and the essential role noise plays in determining the network behavior. In Chapter 2 we present a novel analysis of a previously published dataset of mouse auditory cortex, and extend the current theory of balanced networks to reproduce complicated dynamical behavior. In Chapter 3 we extend the model from chapter 2 to include the study

of noise induced transitions in a bistable neural network. Any actual implementation of a balanced network will naturally deviate away from the theory outlined here, since it is computationally infeasible to simulate an unbounded number of neurons. In particular, the increased variability of balanced networks (as compared to classic rate models) allows for the possibility of nontrivial finite size effects even at reasonably sized implementations of a network. We present a novel result in the study of state transitions and escape times, and provide (what we believe to be) the first work studying a potential landscape that is itself noise dependent.

2.0 INTERLEAVING ASYNCHRONOUS AND SYNCHRONOUS ACTIVITY IN BALANCED CORTICAL NETWORKS WITH SHORT TERM SYNAPTIC DEPRESSION

Internally generated cortical activity is a reflection of the circuit structure and physiology of the cortical network. Circuit models provide an important tool to test and validate whether a specific biological feature of cortex can mechanistically explain recorded population dynamics [68]. However, it is often the case that cortical models are built to capture only a subset of cortical dynamics. Indeed, there are distinct recurrent circuit models for asynchronous population dynamics [69], rhythmic synchrony [70], long timescale dynamics [71], and population-wide coordinated behavior [66, 72]. In many cases these models assume only the circuit structure and physiology needed to replicate the population activity of interest, while ignoring the biology that is critical in the other models. One clear reason for this polychotomy of cortical models is that real cortical dynamics are quite rich, with activity that transitions between distinct states throughout a recording, and models tend to focus only upon a single dynamical state.

For the majority of the time, cortex is in an asynchronous (or weakly correlated) state with temporally irregular spiking dynamics [29, 54, 73, 74] and population responses fluctuate with linear dynamics about an operating point [75, 76]. Models with strong recurrent excitation and inhibition readily capture these dynamics [38, 55, 77], but they require a strong assumption of Gaussian distributed fluctuations in population activity [54]. However, the asynchronous dynamics in biological cortical networks are sporadically interrupted by brief epochs of population-wide coordinated activity, as reported in spontaneously active *in vitro* slice recordings [42, 78–80], spontaneously active cortex in anesthetized animals [45–47], and even in the cortices of awake behaving rodents [81] and primates [82]. These epochs of large

population correlation are clear nonlinear phenomena, with a non-Gaussian character to population activity skewing the activity distribution [37]. Circuit models that capture these population events often do so through a brief respite from short term synaptic depression of recurrent excitation that drives population activity [66, 83–85]. This mechanism ignores the wealth of data suggesting that inhibition confers strong stability to population activity [42, 44, 45, 86–88]. The apparent disconnect between models where population stability is due to strong recurrent inhibition and those where nonlinear runaway population activity emerges through a relief from synaptic depression represents a clear barrier towards a more complete mechanistic theory of cortical dynamics.

In this study we focus on the internally generated dynamics of spontaneously active *in vitro* slices of rodent auditory cortex [42]. Whole cell patch recordings show that the cortical network is largely in a state where recurrent inhibition tracks and balances excitation, except for rare and short periods when the slice undergoes a population-wide surge in activity. (We note that as an aesthetic choice, to suggest excitation as *increasing* the membrane potential and inhibition as *decreasing* it, we have inverted the whole-cell recordings so that excitation (typically shown in red) points *up*, and inhibition (typically shown in blue) points *down*.) A subset of the slice preparations show a population burst of activity with a previously unstudied low frequency (2-12 Hz) within-burst rhythm, while the remaining preparations show population events that lack any rhythm. We present novel analysis and modeling of the complex and varied nature of the population events. To capture this dataset we follow past work [57] and extend balanced network models to include synaptic depression from excitatory neurons to both excitatory and inhibitory neurons. This gives a balanced asynchronous solution where fluctuations in activity may cause a depression-induced weakening of inhibitory recruitment, prompting the excitatory population to explode into a population event. Further, the rhythmic character of population events results from an interplay between recurrent excitation and the depression of excitation. This feature requires that excitatory synapses onto inhibitory neurons exhibit more depression at relatively lower firing rates as compared to depression onto excitatory cells. Fortunately, this requirement has strong support from past electrophysiological recordings in primary auditory cortex [58, 89, 90].

It is clear that the cortex exhibits diverse dynamics poised between asynchrony and

population-wide correlations. Our framework gives a clear blueprint for how to combine and extend past models so as to account for previously unexplained cortical dynamics.

2.1 RESULTS

2.1.1 Two kinds of network stabilization

We first revisit previously published recordings from layer IV neurons in an *in vitro* thalamocortical slice preparation of mouse primary auditory cortex (see [42] for a complete description). An extracellular solution containing high potassium, low magnesium, and low calcium concentrations was used to recruit internally generated neuronal activity within the slice [91]. Whole-cell voltage clamp techniques isolated the excitatory currents into a single neuron by holding its membrane potential at approximately the reversal potential of inhibition (-80 mV). The spontaneous network activity typically provided a net excitation which fluctuated about a low state, suggestive of asynchrony between the pre-synaptic excitatory neurons (Figure 3a). However, this state was sporadically interrupted by large excursions in excitation lasting hundreds of milliseconds, presumably due to coordinated recruitment of excitatory neurons throughout the network (Figure 3a, b). Similar population dynamics have been reported in the spontaneous states of cultured [48, 83], *in vitro* [80, 92], and *in vivo* [81, 93] preparations. We label these rare epochs *population events*, similar in nature to the "population spikes" and "population bursts" described in past studies [66, 84]. In particular, population event dynamics are reported in auditory cortex during *in vivo* spontaneous conditions, determined either directly from population recordings [47] or indirectly from whole cell recordings that show large membrane potential excursions suggestive of widespread synchrony [46].

Population event dynamics are captured by networks of spiking neuron models having recurrent excitatory connections with short term synaptic depression [66, 83–85, 94]. The activity of these models reduces to a simple dynamical system with only excitatory activity and synaptic depression as variables [84, 85, 95, 96], making them amenable to phase plane

analysis (Figure 3c). The positive feedback associated with recurrent excitation could in principle destabilize network activity; however, strong synaptic depression weakens recurrent coupling so that moderate network activity is stable (Figure 3c, intersection of orange and red curves). Consequently, we will throughout refer to these networks as *depression stabilized*. While the asynchronous low activity state is stable, a transient relief from depression promotes a rapid runaway explosion in activity due to strong positive feedback in the excitatory network (Figure 3c, black curve), modeling the population event. This massive excitatory drive recruits synaptic depression that weakens recurrent feedback, ultimately bringing activity back to the low state, terminating the population event.

Dynamical systems with a single stable state that exhibit large scale transient excursions in activity as a response to input perturbations above a certain size are often labeled excitable systems [2]. Population event activity in depression stabilized networks is an example of excitable dynamics, and firing rate model reductions of them require two main assumptions. First, the asynchronous spiking dynamics during the periods between population bursts necessitate either sources of external noise that are independent across neurons [83, 94, 96] or broad neuronal heterogeneities within the population [66, 84]. This ensures that the weak recurrent coupling in the depressed state does not synchronize network activity. Second, synaptic depression must operate on a timescale that is much slower than recurrent excitation [95, 96], otherwise runaway population activity will be prematurely quenched. While the second assumption is well founded in synaptic physiology [58, 64], the first ignores common input projections that would otherwise synchronize the network [76].

In depression stabilized networks inhibitory connectivity plays only a perfunctory role; in some models it is included [66, 83, 85] while in other models it is completely ignored [84, 94–96]. A popular alternative cortical model is one in which strong recurrent inhibition actively counteracts recurrent excitation, effectively stabilizing network activity at a moderate level. Such networks are labeled *inhibition stabilized* [86, 97] and include networks with large excitation that is balanced by recurrent inhibition [38, 54, 55]. Whole cell recording *in vivo* in both visual [86, 88] and auditory cortices [87] have provided strong evidence for an inhibition stabilized cortex.

Depression and inhibition stabilized networks differ in several key aspects. In inhibition

stabilized networks the timescale of inhibition must be sufficiently fast to prevent runaway activity. Because of this, inhibition stabilized networks cannot produce a stereotypical population event, and rather produce only stunted transient responses to a slight reduction in inhibition (Figure 3d). Nevertheless, networks with strong recurrent inhibition require neither sources of independent noise external to the network nor broad cellular heterogeneities to produce asynchronous dynamics; rather, they produce it naturally through inhibitory cancellation of correlations due to common excitation [54, 77].

One advantage of spontaneously active *in vitro* slice preparations is that joint excitatory-inhibitory dynamics can be directly measured [42]. Simultaneous whole-cell recordings from neuron pairs were obtained, where one neuron's membrane potential was held at the inhibitory reversal potential (-80 mV; Figure 3e, red curve) and the other at the excitatory reversal potential (0 mV; Figure 3e, blue curve), isolating excitatory and inhibitory population activity, respectively. Restricting analysis of paired neuron activity to the asynchronous period (i.e. excluding population events) shows that while neuron pairs receive both correlated excitation (Figure 3f, $E \rightarrow E$) and inhibition (Figure 3f, $I \rightarrow I$), the composite input to the pair measured at resting potential is only weakly correlated (Figure 3f, Total). This suggests that shared inhibition actively tracks and cancels shared excitation, producing an asynchronous network state as has been previously theorized [54, 77], and is a hallmark of inhibitory stabilization.

Taken together these observations present a problem when considering the spontaneously active *in vitro* auditory cortex slice data (Figure 3; [42]). Specifically, the excitatory-inhibitory dynamics recorded between population events support inhibitory stabilized dynamics. Such a cortical model disallows the existence of population event dynamics, and yet population events are clearly present in the data and inhibition is strongly recruited during the excursion (Figure 3e). The central focus of our study is to provide a circuit model of cortex that captures the full range of asynchronous and population event dynamics in auditory cortex.

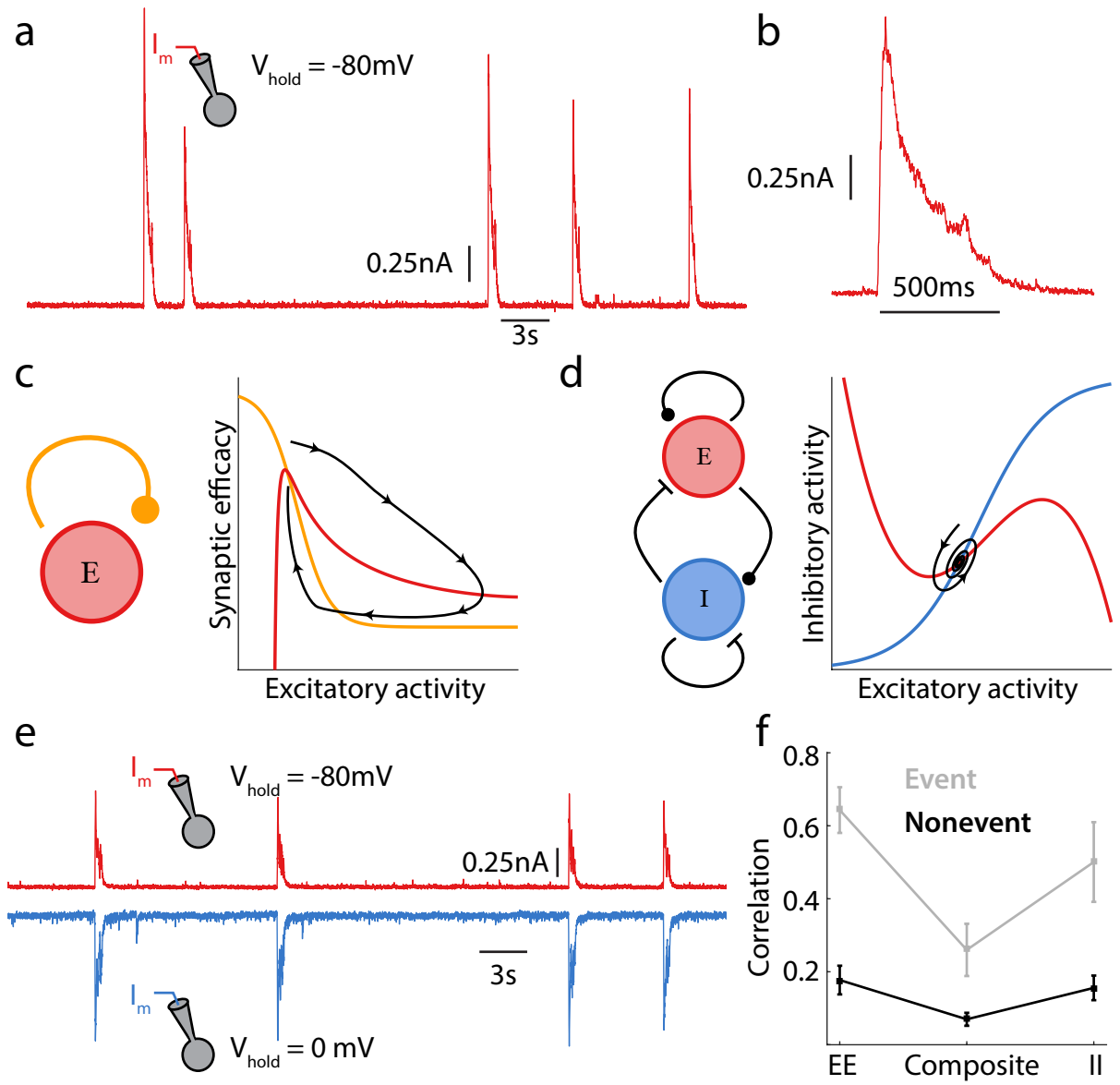

Figure 3: Two mechanisms of stabilization. **a:** Example whole cell voltage clamp recording from mouse auditory cortex, *in vitro*; the neuron is held at the reversal potential for inhibition. Data inverted. **b:** Magnification of population event from recording shown in panel a. **c:** Phase portrait for a depression stabilized network; the red and orange curves denote nullclines for excitation and synaptic efficacy (depression), respectively. A momentary release of depression permits large runaway excitatory activity until more depression is recruited, decreasing excitatory activity (black curve). **d:** Phase portrait for an inhibitory stabilized network; the red and blue curves denote nullclines for excitation and inhibition, respectively.

A perturbation to network activity causes a brief and small amplitude relaxation back to equilibrium (black curve).**e:** Synaptic input traces from dual whole cell patch clamp recordings from mouse auditory cortex. Two neurons are simultaneously clamped, one held at the reversal potential of excitation, the other at the reversal potential of inhibition. Data inverted. **f:** The correlation coefficient of synaptic inputs from simultaneously recorded pyramidal neurons. The neuron pair is held at the inhibitory/inhibitory reversal to measure the correlation of excitatory currents (EE), or held at the excitatory/excitatory reversal to measure the correlation of inhibitory currents (II), or held at a membrane potential where both excitatory and inhibitory currents are active to measure the correlation of summed excitation and inhibition (composite). Analysis is shown for current fluctuations during population event and non-event periods.

2.1.2 Oscillatory and non-oscillatory population events

In order to better motivate our model we first analyze the structure of population activity within events recorded from our spontaneously active cortical slice [42]. We note that any arbitrary pair of recorded cells likely do not share a direct connection, though all patched cells register simultaneous population events throughout the recording. We take this as an indication that the population event is slice-wide, and not likely an artifact of recording site.

Broadly speaking, there are two classes of population events – ones where the synaptic excursion is a simple rise and then decay (Figure 4a, lavender curve) and ones with a low frequency rhythmic dynamic throughout the population event (Figure 4a, green curve). Even so, population events display a wide range of heterogeneity across all slices (Figure 4a, black curves). The population events recorded from a given neuron would not switch between these classes: if a neuron showed a rhythmic event during a recording, all events recorded from that neuron would be rhythmic. This is verified by spectral analysis restricted to the population event epochs from these two example neurons, showing that the respective arrhythmic or rhythmic character was recording site specific (Figure 4b; see Methods). Spectral analysis of the population event epochs across the entire dataset ($n = 210$ from 23 distinct slices) shows that approximately one third of recorded neurons exhibited rhythmic events (Figure 4c). In

the oscillatory cases the frequency is relatively slow, and somewhat heterogeneous across the slices, ranging from $\sim$ 2Hz to 12Hz (Figure 4d). A natural question to ask is whether population event dynamics are due to a synaptic and cellular property of the recorded neuron, or a network-wide feature.

Reordering the dataset according to slice membership rather than spectral width shows a slice dependent clustering of oscillatory event dynamics (Figure 4e). This is consistent with an interpretation of a population event as a network feature, since if one neuron experiences an oscillatory event then all other neurons in the same slice do as well. Further, the frequency of the oscillation is also highly correlated within a slice; namely, the peaks in the power spectrum for simultaneously recorded neurons lie roughly on the identity line (Figure 4f). Taken together, these results suggest that the oscillatory character of population events is due to slice heterogeneity, rather than a cellular process that is private to a neuron. This fact will simplify the assumptions of our cortical model.

In contrast to the population event dynamics, the synaptic fluctuations that occurred between population events were dominated by low frequency power, and showed no discernible peak suggestive of rhythmic dynamics (Figure 4g). A lack of pronounced population-wide oscillations is characteristic of the asynchronous state associated with the low activity regime [29, 54, 77]. The fact that the third of slices which show oscillatory dynamics during the event do not show it for the low activity state suggests a strong nonlinearity in network dynamics. Overall, these data provide an important constraint – our circuit model must be able to produce both arrhythmic and rhythmic population events, as well as show a clear asynchronous low activity state devoid of any rhythm between population events.

2.1.3 Balanced cortical circuit model of population events

We first begin with modeling the low activity asynchronous state in which the cortical population resides for most of the time. Motivated by the fact that recurrent inhibition is critical for asynchrony in the auditory cortical slice recordings (Figure 3f), we consider recurrently coupled populations of excitatory (E) and inhibitory (I) neurons (Figure 5a). The excitatory and inhibitory populations have N_E and N_I neurons, respectively. For simplicity we

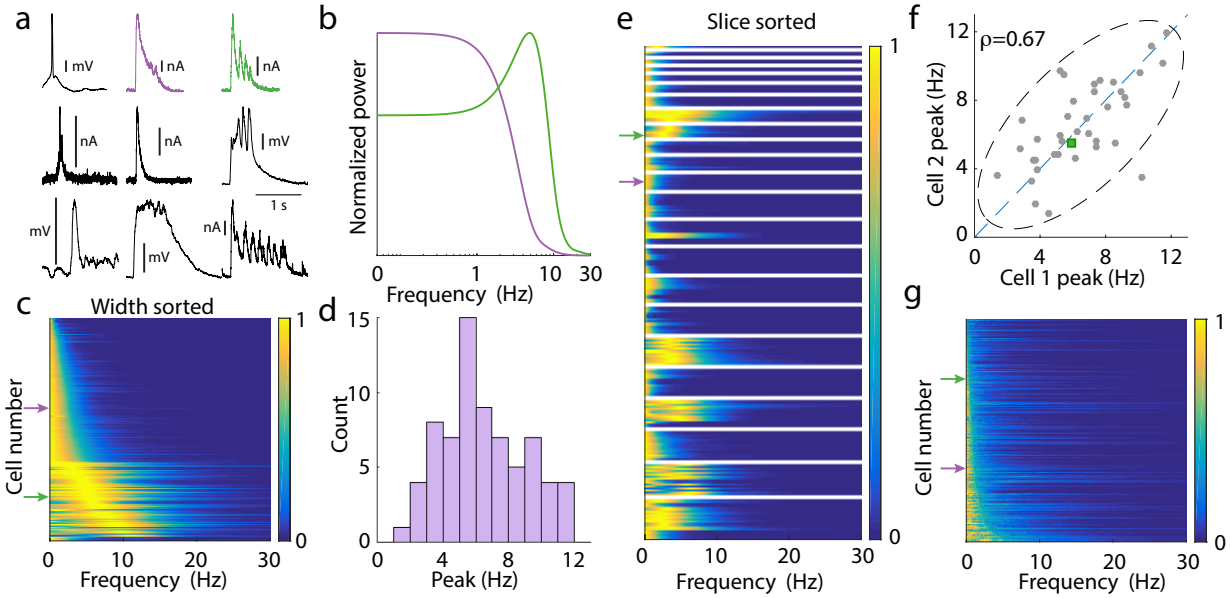

Figure 4: Heterogeneity of the dynamics within population events. **a:** Sample population events from different slices. For events recorded in voltage clamp, vertical scale bar denotes 0.2nA. For events recorded in current clamp, vertical scale bar denotes 10mV. Sample events marked in lavender (aperiodic) and green (periodic), are exemplar events marked throughout figure. All sample traces inverted to point up for consistency. **b:** Power spectra for marked cells from the dataset to the left, averaged over all of the events seen. There is a clear peak in the power spectrum, here at $\sim$ 6Hz. Power spectra are normalized to have maximum value 1. **c:** The power spectra for the event periods of the entire dataset. Power spectra are normalized to have maximum value 1. Data sorted by full width at half max (for those without peaks), and then by peak location (for those with peaks). Arrows denote marked data from panel *a*. **d:** Histogram of the peak location for peaked power spectra. **e:** Same data as panel *c*, sorted and grouped by slice, rather than by width. **f:** Scatterplot of the peak location for pairs of cells. Since some of the collected data measures excitatory currents, some measures inhibitory currents, and some measures mixtures, there is no natural categorization for choosing which cell is "Cell 1" and which is "Cell 2". These labels are chosen at random for this figure. The reported correlation coefficient is the median of the distribution of correlation coefficients computed for different instantiations of the randomization. Dashed ellipse denotes the 95% confidence interval. Periodic data from panel *a* marked as a green square. **g:** Same as panel *c* for the nonevent portions of the data.

take $N_E = N_I \equiv N$, but this assumption can be easily relaxed. Since we will be studying network-based mechanisms we opt for a phenomenological network model that only considers neurons as being active or inactive. We focus on the proportion of active neurons in population α, $r_\alpha(t) = n_\alpha(t)/N$, where $n_\alpha(t)$ is the number of active neurons at time t (here $\alpha \in \{E, I\}$).

Following classical work in the theory of balanced excitation and inhibition [54, 55] we scale the baseline synaptic strength from a neuron in population β to a neuron in population α as $J_{\alpha\beta} = j_{\alpha\beta}/\sqrt{K}$, where $j_{\alpha\beta} \sim \mathcal{O}(1)$ is the unscaled connection strength ($\alpha, \beta \in \{E, I\}$), and K is the average number of connections onto a neuron. Here we will take $K = cN$, with $c \in (0, 1)$ being a fixed and positive number. This scaling for $J_{\alpha\beta}$ implies very strong connections between neurons, especially when compared to those from the more common $1/K$ scaling. As a result the network produces sizable internally generated variability, so that single neuron activity is temporally irregular, in line with experimental recordings [73]. While $1/\sqrt{K}$ synaptic scaling was originally a theoretical abstraction, recent work in cultured neuronal preparations gives strong evidence for this scaling [48].

We take a sigmoidal activation function f_α for population α, and for large K the population activity then obeys the following dynamics (see Methods):

$$\begin{aligned} \dot{r}_{_E} &= -r_{_E} + f_E\Big(\sqrt{K}\underbrace{(j_{EE}p_{EE}r_{_E} - j_{EI}r_{_I} + I_{_E})}_{\text{Balance condition: } \mathcal{O}(1/\sqrt{K})}\Big), \\ \tau_{_I}\dot{r}_{_I} &= -r_{_I} + f_I\Big(\sqrt{K}\underbrace{(j_{IE}p_{IE}r_{_E} - j_{II}r_{_I} + I_{_I})}_{\text{Balance condition: } \mathcal{O}(1/\sqrt{K})}\Big). \end{aligned} \tag{2.1}$$

Here $\tau_{_I}$ is the time constant of the inhibitory population relative to the excitatory, and I_α is a source of external drive to population α. Finally, $p_{\alpha E}$ quantifies the degree of short term synaptic depression from the E population onto population α. To gain intuition for the subtleties of this scaling, for the moment we ignore depression and set $p_{\alpha E} = 1$. The $1/\sqrt{K}$ synaptic scaling introduces a complication, namely that the argument to f_α in principle grows with $\sqrt{K}$. This could produce saturated activity for large K, disallowing any reasonable cortical dynamics. However, the network robustly corrects for this through a *balance condition* whereby the combined external and recurrent excitation is roughly can-

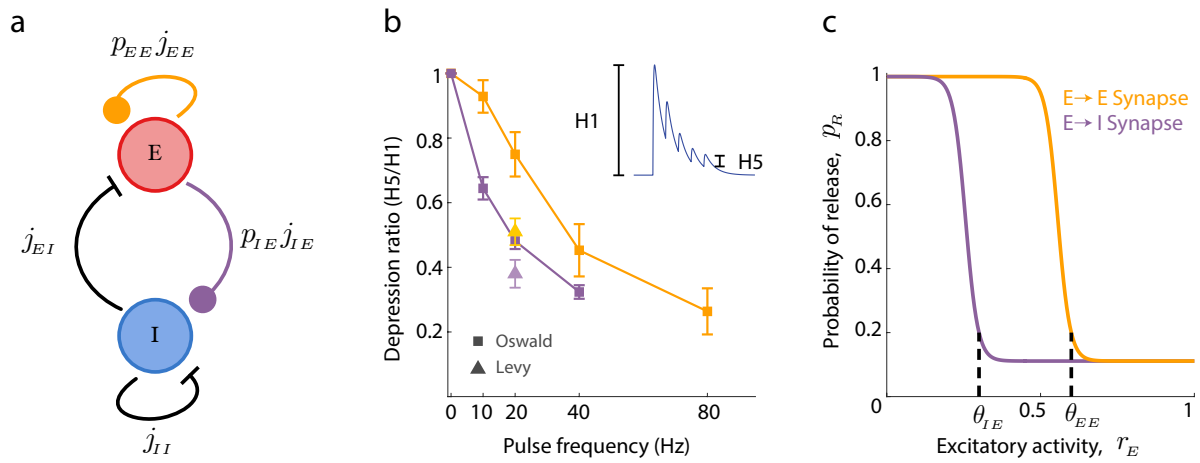

Figure 5: Balanced network model with synaptic depression of excitatory connections. **a:** Schematic of the full model. Excitatory (red) and inhibitory (blue) populations are modeled as Markov processes, while the dynamical synapse variables (p_{EE}, p_{IE}) are modeled as continuous, slowly varying variables. Connection strengths ($j_{\alpha\beta}$) are marked. Dynamical variables are all scaled to live in $[0, 1]$. **b:** Data from Oswald & Reyes [89] and Levy & Reyes[98] showing amount of depression as a function of pulse frequency. Depression is measured as the ratio of the response heights from the fifth spike to the first spike (see *Inset*). Synapses between pyramidal cells are shown in orange, synapses from pyramidal to inhibitory shown in purple. **c:** Release probability (p_R) as a function of excitatory firing rate (r_E) used in the model. Threshold parameters $\theta_{\alpha\beta}$ are marked. The only difference between the curves is the threshold for onset of depression. To agree with the data in panel *b*, the threshold for the EE synapse is chosen as larger than the threshold for the IE synapse ($\theta_{EE} > \theta_{IE}$).

celed by recurrent inhibition so that the total input remains $\mathcal{O}(1)$ for all network sizes. For robust balance conditions, one requires large external drive, i.e. $I_\alpha \sim \mathcal{O}(1)$. Such a balance condition produces asynchronous network dynamics [54], and the dual patch whole cell experiments in [42] (Figure 3) were originally performed as a validation of this theory.

While balanced network models have been successful at explaining several aspects of cortical spiking dynamics, for $K \to \infty$ the firing rates are nonetheless determined by the pair of linear equations defined by the balanced conditions. In particular, this means that the firing rates (r_E, r_I) are linearly related to the external inputs (I_E, I_I), and the nonlinearities of the transfer functions f_α do not contribute to population activity. Additionally, the timescale of cortical dynamics becomes very fast, so that for moderate τ_I there cannot be rhythmic dynamics [55]. These limitations are serious obstacles when trying to capture the clearly nonlinear population event behavior (Figure 3), as well as the slow rhythmic dynamics in a subset of the population events (Figure 4).

Mongillo et al.[57] have extended the theory of balanced networks to include short term synaptic plasticity. Following their example, we use previous mean field reductions [64, 67] to model population-wide synaptic depression, $p_{\alpha E}$, from the E population onto population α:

$$\frac{dp_{\alpha E}}{dt} = \frac{1 - p_{\alpha E}}{\tau_\alpha^r} - \frac{a_\alpha(r_E) p_{\alpha E}}{\tau_\alpha^d}, \tag{2.2}$$

$$a_\alpha(r_E) = \frac{m_\alpha}{1 + e^{-\beta_\alpha (r_E - \theta_{\alpha E})}}. \tag{2.3}$$

The first term on the right hand side of Eq. (2.2) models the recovery from synaptic depression while the second term models its recruitment; these occur on timescales τ_α^r and τ_α^d, respectively. While recovery dynamics are simple, recruitment depends nonlinearly on the E population activation through $a_\alpha(r_E)$. This recruitment is sigmoidal in r_E (Eq. (2.3)), where $\theta_{\alpha E}$ is the threshold level in r_E required for significant depression in the excitatory synapse onto neurons in population α. When coupled with the balance conditions in Eq. (2.1), the short term depression dynamics in Eq. (2.2) and (2.3) allow the nonlinearities in a_α to shape the firing rate solutions (r_E, r_I). This is because synaptic depression now reduces the baseline connectivity $j_{\alpha E}$ by an r_E-dependent solution of Eq. (2.2) and (2.3). Thus the network firing

rates (r_E, r_I) must be solved in conjunction with the equilibrium solution of $p_{\alpha E}$, involving strong nonlinearities. For instance, in the original Mongillo et al. study [57], short term facilitation of excitatory-excitatory connectivity produced a bistable network with coexisting low and high activity regimes.

Short term depression involves several complicated chemical and biophysical processes [59], all lending to significant heterogeneity of plasticity recruitment and recovery across synapse types [99]. In layer II-III of auditory cortex the short term depression of both the excitatory to inhibitory and excitatory to excitatory synapses has been measured through simultaneous *in vitro* patch clamp recording [89, 90] (the results are summarized in Figure 5b). Here, it is clear that $\theta_{IE} < \theta_{EE}$, meaning that the recruitment of depression of excitation onto the inhibitory population occurs at lower r_E than depression onto the excitatory population (compare purple vs. orange curves in Figure 5b). Similar recordings from layer IV of auditory cortex [98] are consistent with the overall trend we observe (Figure 5b, triangles). We adopt this depression heterogeneity in our model (Figure 5c), and we will show this to be a critical requirement to capture the full range of the population behavior recorded in the auditory cortical slice. We remark that for simplicity our model ignores any short term plasticity of the inhibitory connectivity, despite these connections being shown to depress in auditory cortex [90, 100].

In the large network size limit ($K \to \infty$) the dynamics of Eq. (2.1) to (2.3) lack any stochastic aspect to their population activity. This means that solutions cannot capture the random activation time of population bursts that is a clear feature of the *in vitro* data. The collective behavior of finitely many neurons often introduces new stochastic effects at the population level that are absent in theories that take $K \to \infty$ [29, 101–105]. To explore our model with finite K we study the associated birth-death Markov process whose large K limit is the circuit model in Eq. (2.1) [103]. Briefly, the number of active neurons n_α undergoes stochastic transitions to either $n_\alpha + 1$ (birth) or $n_\alpha - 1$ (death) (Figure 6a; see Methods). The stochastic birth and death of population activity imparts a random character to $r_\alpha(t)$ that is shaped by the recurrent excitatory/inhibitory circuit dynamics along with the slower synaptic depression. Simulations of the finite size cortical network show fluctuating population activity about a low state, punctuated by sporadic population events (Figure 6b,c,

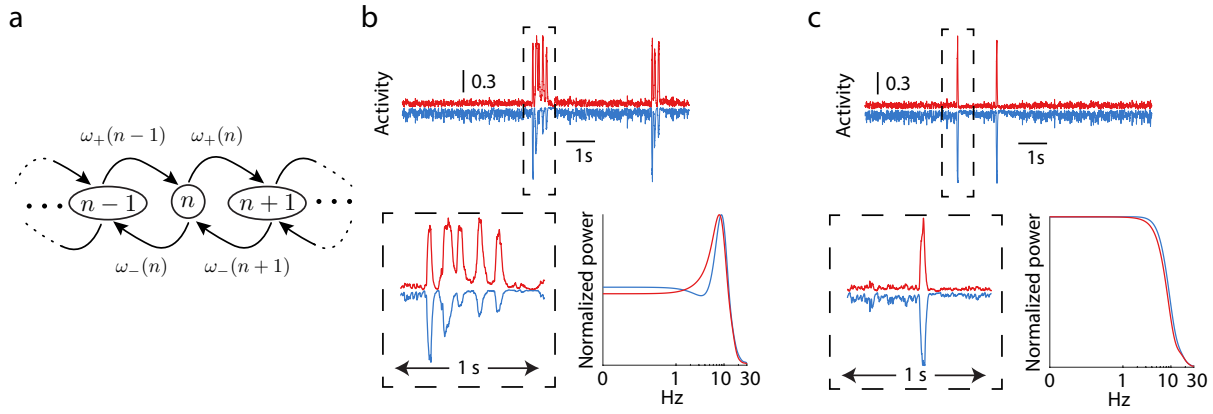

Figure 6: Birth-death Markov model of network activity. **a:** Each neuron is modeled as a binary neuron that can be in an "active" or "inactive" state, and we measure the number of neurons n in the "active" state. The transition rates between states, ω_+ and ω_-, are shown. **b-c:** Simulations for two different parameter sets that reproduce the rare, random nature of the events seen in data. In both panels, $K = 400$. *b:* These events display an oscillatory nature. A simulation of 10s is show across the top, while a 1s interval around the first event is shown on the bottom left. Power spectra for the excitatory and inhibitory populations during the event is shown on the bottom right. This simulation shows an oscillation at ~ 8 Hz. Power spectra are normalized to have maximum value 1. Here $\theta_{EE} = 0.2, \theta_{IE} = 0.5$. *c:* Same as *B*, except these events display no oscillatory nature. The power spectrum shows no clear peak, but does have a large degree of low frequency power. Here $\theta_{EE} = 0.8, \theta_{IE} = 0.5$.

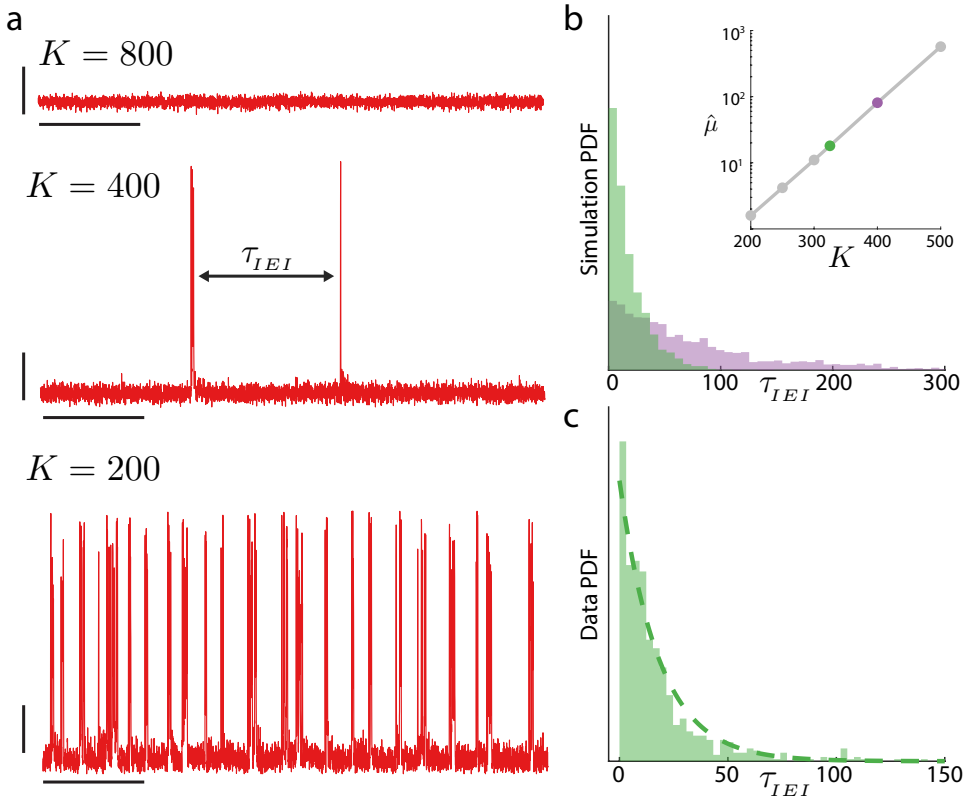

Figure 7: Noise due to finite network size induces population events. **a:** Sample simulation trajectories of the excitatory population activity for three different system sizes K. Vertical scale bars indicate 0.1 a.u., and horizontal scale bars indicate 10s. Each trace shows a total of 50s of simulation. As K increases, events become exponentially less likely. τ_{IEI} is the time between two successive events. **b:** τ_{IEI} is exponentially distributed, consistent with a noise-induced escape model. $\hat{\mu}$ is the exponential parameter fit from a maximum likelihood estimate (MLE) of simulated data; as expected $\hat{\mu} \sim \exp(K)$. Highlighted are distributions for $K = 325$ (green) and $K = 400$ (purple). Connection strengths between populations were scaled for these simulations to fix the underlying vector field to isolate the system-size/noise relationship. Histograms are scaled to have total area 1. **c:** Distribution of inter-event times from recorded *in vitro* data (Figure 3 and 4). The distribution appears exponentially distributed, despite being aggregated from all slices. Dashed line is the MLE fit to the data, giving approximately the same $\hat{\mu}$ as the simulation for $K = 325$ (Data: $\hat{\mu} = 18.03$, Sim: $\hat{\mu} = 18.05$).

top). Depending on the depression thresholds θ_{EE} and θ_{IE} the population events show either rhythmic (Figure 6b, bottom) or arrhythmic (Figure 6c, bottom) dynamics.

The statistics of population events from our model behaves as expected for rare processes. In the $K \to \infty$ limit population-wide fluctuations vanish and consequently population events will never occur. As K decreases, fluctuations increase and thus events become more and more likely, until finally for low enough K the finite size network has behavior significantly different from the large K limit (Figure 7a). A hallmark of rare noise induced transitions is an exponentially distributed inter-event interval[101]. Over a large range of K our model agrees (Figure 7b), and the exponential parameter grows exponentially with system size (Figure 7b inset), again in agreement with rare event theory due to finite system size [29, 101–105]. Satisfyingly, exponentially distributed inter-event intervals are found in our *in vitro* data, aggregated over all events and all slices (Figure 7c). Moreover, qualitative agreement in inter-event interval distribution is found between the model and the data for reasonably sized networks (K on the order of a few hundred). This is in contrast to the traditional $1/K$ scaling, where rare events are often observed only for exceptionally small networks, with K on the order of tens. In total, the circuit model in Eq. (2.1) to (2.3) recapitulates the main features of the asynchronous and population event epochs of the spontaneously active cortical slice experiments.

The simplicity of the model makes it amenable to analysis that reveals how the recurrent excitation and inhibition combine with the two synaptic depression processes to produce rich population event dynamics. In the following sections we will use the fact that the $E \to I$ synapses depress at lower pre-synaptic rates than $E \to E$ synapses ($\theta_{IE} < \theta_{EE}$; Figure 5b,c) to probe the separate mechanisms underlying population *event initiation* and *event rhythmicity*.

2.1.4 Population event initiation

Population event initiation occurs at r_E values that are far below θ_{EE}, and thus a reasonable approximation for the low activity regime is to take the $E \to E$ synapses as undepressed ($p_{EE} = 1$) (Figure 8a). This simplification reduces the network dynamics in Eq. (2.1) to (2.3) to three dynamical variables: $r_E(t)$, $r_I(t)$, and $p_{IE}(t)$. In this reduced system we consider

how population activity depends on the threshold for $E \rightarrow I$ synaptic depression, θ_{IE}. For sufficiently low θ_{IE}, so that the $E \rightarrow I$ synapse is depressed even at low r_E values, the excitatory population cannot recruit sufficient recurrent inhibition to stabilize a low activity state. Rather, only a stable high activity state exists (Figure 8b, solid line at $r_E \approx 1$). In this regime an external (negative) input can transiently lower $r_E(t)$, allowing $p_{IE}(t)$ to recover and $r_I(t)$ to increase despite the reduction in r_E (Figure 8c, left, 3s $< t <$ 5s). Nevertheless, after stimulus removal the network quickly returns to the high activity state with $p_{IE}(t)$ returning to very low values (Figure 8c, left, $t >$ 5s). Thus, cortical models with small θ_{IE} that do not allow a stable low activity state cannot hope to capture the *in vitro* population activity (Figure 3a).

As θ_{IE} increases a stable low activity state is created (Figure 8b, solid line at $0 < r_E << 1$) along with an unstable threshold state (a saddle point) that separates the new low activity regime from the high activity regime (Figure 8b, dashed line). Here the dynamics are bistable; when $r_E(t)$ is in the low activity state a sufficiently strong input can transition the network into the high active state (Figure 8c, middle, 3s $< t <$ 5s). The transition recruits sufficient $p_{IE}(t)$ depression so that the network remains in the high state after the stimulus is removed (Figure 8c, middle, $t >$ 5s). This transition from the low to high activity state of r_E is the population event initiation in the full model. Finally, for very large θ_{IE} the network remains bistable, but the saddle point is distant from the low activity state (Figure 8b). Here, moderate perturbations no longer promote transitions from low to high population activity (Figure 8c, right). For finite size networks internally generated fluctuations can initiate a population event via a stochastic transition across the stable manifold of the saddle point; these transitions are far more likely for moderate θ_{IE} where the stable low state and saddle points are near one another.

In sum, the reduced model shows how an activity dependent weakening of inhibitory recruitment allows a complex dynamic, wherein a low activity inhibitory stabilized state can have a threshold beyond which inhibition can no longer prevent runaway activity from positive feedback. However, our reduced model (with $p_{EE} = 1$) lacks the requisite mechanisms to produce excursions of stereotyped width that are clear in the population events of the full model (Figure 6b,c).

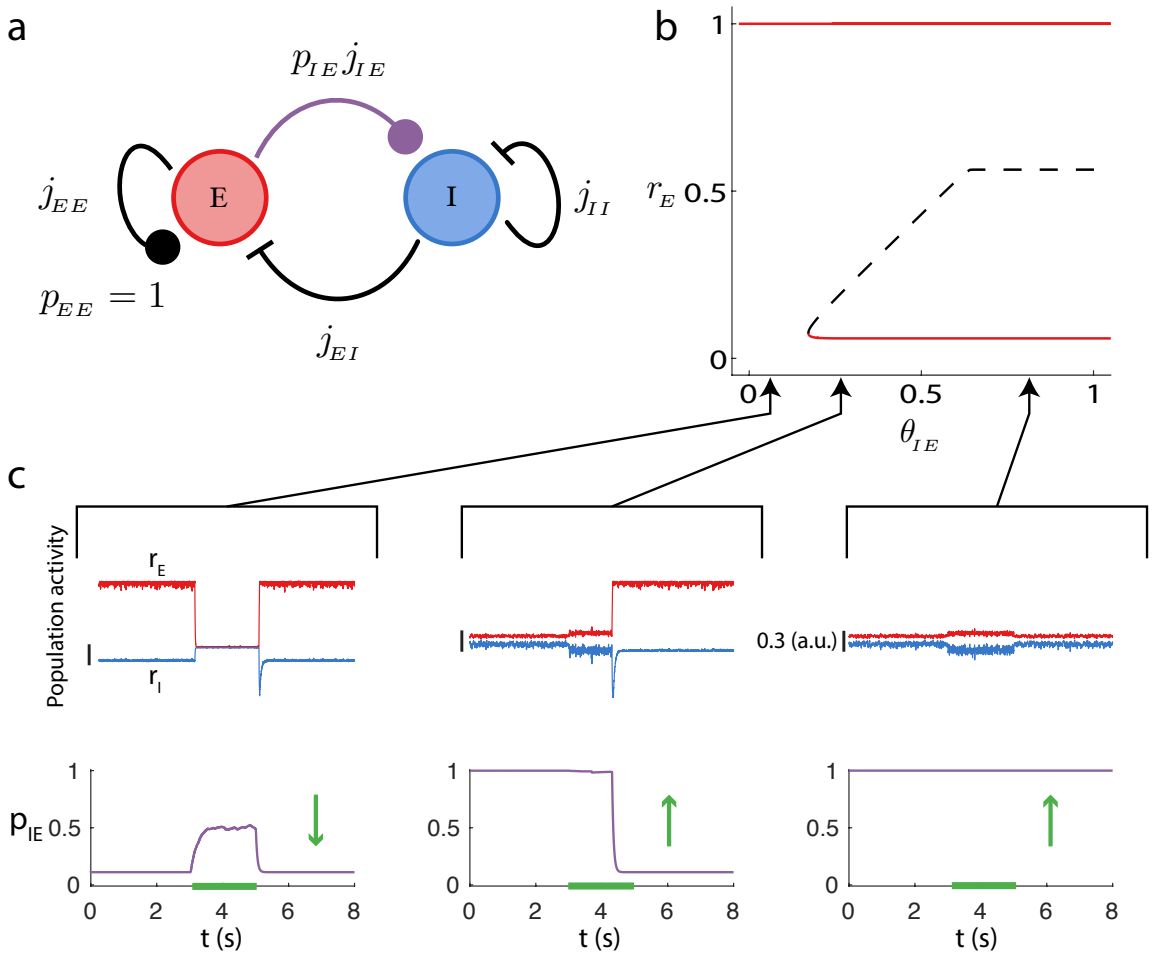

Figure 8: Reduced network model with fixed EE synaptic strength in the undepressed state ($p_{EE} = 1$). **a:** Schematic of the reduced system. **b:** Bifurcation diagram of the excitatory rate (r_E) as the depression threshold θ_{IE} is varied. Red lines denote the stable fixed points and the dashed black line denotes an unstable (saddle) fixed point. The lower stable state and saddle point coalesce in a saddle-node bifurcation is the balanced state (intersection of lower red and dashed lines). **c:** Example simulations for three values of θ_{IE}, with $K = 400$. Top row: traces of the activity, r_E in red, r_I in blue. Vertical scale bars denote 0.3, arbitrary units. Bottom row: plasticity variable. We applied an external stimulation to the excitatory population from $t = 3$s to $t = 5$s (green bar marks external stimulation). The depression threshold θ_{IE} is indicated by the black arrows linking to panel b.

2.1.5 Population event rhythmicity

To investigate the mechanisms underlying within population event dynamics we study a reduced model with the $E \to I$ synapses assumed to be depressed ($p_{IE} \approx 0.11$, Figure 9a), as is the case after event initiation (Figure 8). This reduction restricts the network dynamics in Eq. (2.1) to (2.3) to three dynamical variables: $r_E(t)$, $r_I(t)$, and $p_{EE}(t)$. In particular, a low p_{IE} value removes the low activity state, and the reduced model is then only appropriate for modeling network dynamics within a population event epoch.

In this reduced model a large threshold for synaptic depression in the $E \to E$ synapse, θ_{EE}, produces a stable activity state at high r_E (Figure 9b, solid red curve and top inset). However, as θ_{EE} is reduced the recurrent excitation is depressed and the high state loses stability via a super-critical Andronov-Hopf bifurcation. Consequently, this births a stable limit cycle solution of growing amplitude, creating oscillatory dynamics within the network (Figure 9b, green curves and bottom inset). The oscillation reflects a competitive interplay between $p_{EE}(t)$ and $r_E(t)$, where the oscillatory timescale is set primarily by the depression recruitment and recovery time constants. Further, over a wide range of θ_{EE} the frequency is between 4 and 8 Hz (Figure 9b, bottom), matching the oscillation frequency in the full model (Figure 6b, bottom) and a subset of the *in vitro* data (Figure 4c). Overlaying the limit cycle on the plasticity nullcline in the (p_{EE}, r_E) plane shows how the oscillation samples the full nonlinearity of p_{EE} recruitment (Figure 9c, cf. Figure 3c), highlighting the nontrivial interaction between p_{EE} and r_E.

The combination of the two reduced models (Figure 8 and Figure 9) suggests that the two depression pathways, p_{IE} and p_{EE}, play distinct roles in population event initiation and within event dynamics, respectively. Armed with these insights we next aim to study how population event dynamics depends upon the pair of depression thresholds (θ_{EE}, θ_{IE}) in the full model.

2.1.6 Suite of population event dynamics

To study the range of possible dynamics in the full model we begin by constructing a two parameter (θ_{EE}, θ_{IE}) bifurcation diagram of Eq. (2.1) to (2.3) (center panel of Figure 10). There are two main organizing structures in the diagram. First, a curve of saddle-node bifur-

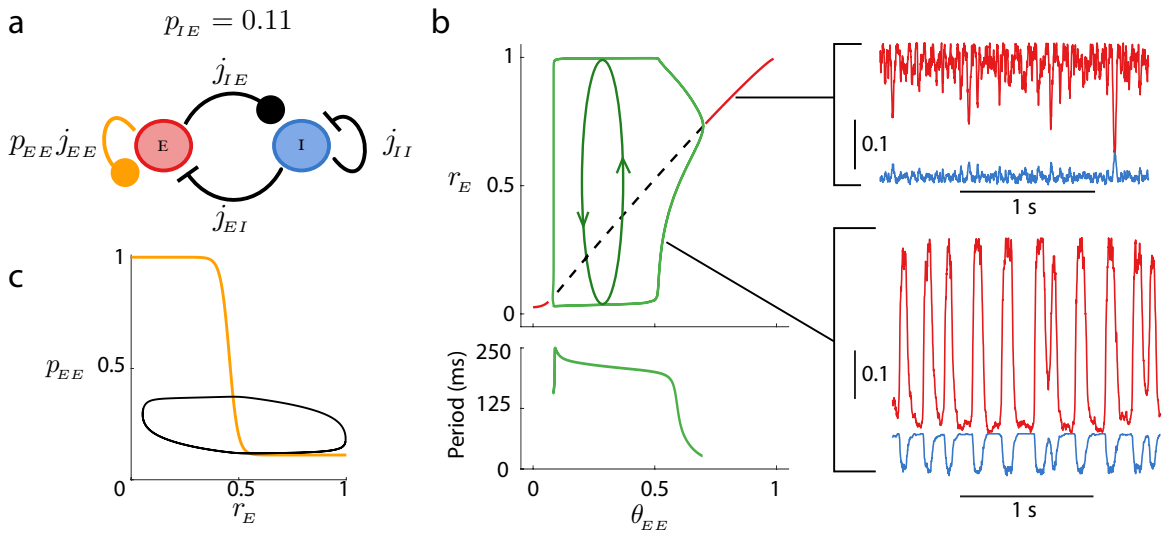

Figure 9: Reduced network model with fixed EI synaptic strength in the depressed state ($p_{IE} = 0.11$). **a:** Schematic of the reduced system. **b:** *Top:* Bifurcation diagram of the excitatory rate (r_E) as the depression threshold θ_{EE} is varied. Red lines denote stable fixed points, the dashed black line denotes an unstable fixed point, and the green lines show the peak and trough of a network oscillation (limit cyle). *Bottom:* The period of the oscillation as a function of θ_{EE}. *Right:* Sample simulations for two different values of θ_{EE} (top: $\theta_{EE} = 0.9$, bottom: $\theta_{EE} = 0.5$), with $K = 400$. **c:** Limit cycle of the underlying dynamical system overlaid on the nullcline for p_{EE}.

cations (purple curve in Figure 10, cf. Figure 8b) separates regimes with either monostable or bistable dynamics. Bistable dynamics exist for sufficiently large θ_{EE} and moderate θ_{IE} (purple region in Figure 10), while either a low rate solution (yellow region in Figure 10) or saturated state (gray region in Figure 10) are possible outside of this region. Second, curves of Andronov-Hopf bifurcations (green curves in Figure 10, cf. Figure 9b) demarcate a region with oscillatory dynamics (green region in Figure 10). The combination of saddle-node and Andronov-Hopf bifurcations provides key intuition into how the network responds to finite K induced internally generated fluctuations.

For sufficiently large θ_{IE} (so that the system is above both the saddle-node or Andronov-Hopf bifurcations), there is only a stable low state (yellow region in Figure 10). For low θ_{EE} the recurrent excitation is very depressed and internally generated fluctuations simply perturb the activity about the low state (upper left inset in Figure 10). Moderate θ_{EE} makes it possible for fluctuations to transiently remove depression, resulting in a large population event that is akin to the excitable dynamics explored in past studies [66, 83–85, 96] (top inset in Figure 10). However, because the network has not passed through the Andronov-Hopf bifurcation it lacks any oscillatory dynamics and consequently these population events are strictly arrhythmic.

For very low θ_{IE} and moderate θ_{EE} the system has passed through the Andronov-Hopf bifurcation and a stable limit cycle solution produces periodic dynamics (bottom left inset in Figure 10). However, for larger θ_{IE} the system passes through the saddle-node bifurcation and more subtle population dynamics emerge (dark green region in Figure 10). In this region there is no longer a stable limit cycle to organize periodic behavior; rather, there is only a stable low rate state. Nonetheless, finite size fluctuations induce transient population bursts with oscillatory dynamics (bottom inset in Figure 10). This region of parameter space contains the $(\theta_{EE}, \theta_{IE})$ pairings for the initial simulations of the full model (Figure 6).

Focusing on $(\theta_{EE}, \theta_{IE})$ pairings within this region we find that it is possible to have oscillatory and non-oscillatory population event dynamics. For very low θ_{IE} (point I in Figure 11a) there is a globally stable limit cycle solution and an unstable fixed point (phase portrait I in Figure 11b). Simulations of the full finite size stochastic model show a robust and sustained low frequency oscillation (time series I in Figure 11b). As θ_{IE} is increased the

system passes through a period doubling bifurcation (red curve in Figure 11a). This shifts the stable limit cycle to have period two dynamics (II in Figure 11b); we remark that the limit cycle remains globally stable. As θ_{IE} is increased further another period doubling bifurcation (orange curve in Figure 11a) gives rise to stable period four dynamics (not shown). As θ_{IE} continues to increase, the system undergoes a period doubling cascade,
suggesting complicated chaotic transient dynamics. For even larger θ_{IE} a stable/unstable pair of fixed points are born out of the saddle-node bifurcation (purple curve in Figure 11a), and the stable limit cycle is lost. Here the full model dynamics is excitable, with a slight perturbation from the globally stable fixed point resulting in a large scale excursion before returning to the fixed point (phase portrait III in Figure 11b). However, the high period structure of the stable cycle that existed below the saddle-node bifurcation still shapes the dynamical flow, so that the excursion has a rhythmic character. This is apparent in the fluctuation-induced population event dynamic (time series III in Figure 11b). Finally, for even larger θ_{IE} the excitable dynamics persists, but now the excursion is simple and the population event lacks rhythmic dynamics (IV in Figure 11b). This is because we are now farther in $(\theta_{EE}, \theta_{IE})$ parameter space from the region where the stable high period limit cycle shaped population dynamics. These last two parameter sets (III and IV in Figure 11) capture that rhythmic and arrhythmic population event dynamics, respectively; indeed spectral analysis of the time series from the stochastic model (Figure 11c) reflect that from the *in vitro* data (Figure 4b).

2.2 DISCUSSION

In this work we studied the population dynamics of finite size balanced networks with short term depression of excitatory projections. Despite the model's simplicity it provided a parsimonious explanation of recordings from spontaneously active *in vitro* slices of auditory cortex [42]. In particular, population events were sporadic and random in time, and depending on short term plasticity thresholds the event could have rhythmic or arrhythmic dynamics, all in agreement with analysis of the cortical recordings.

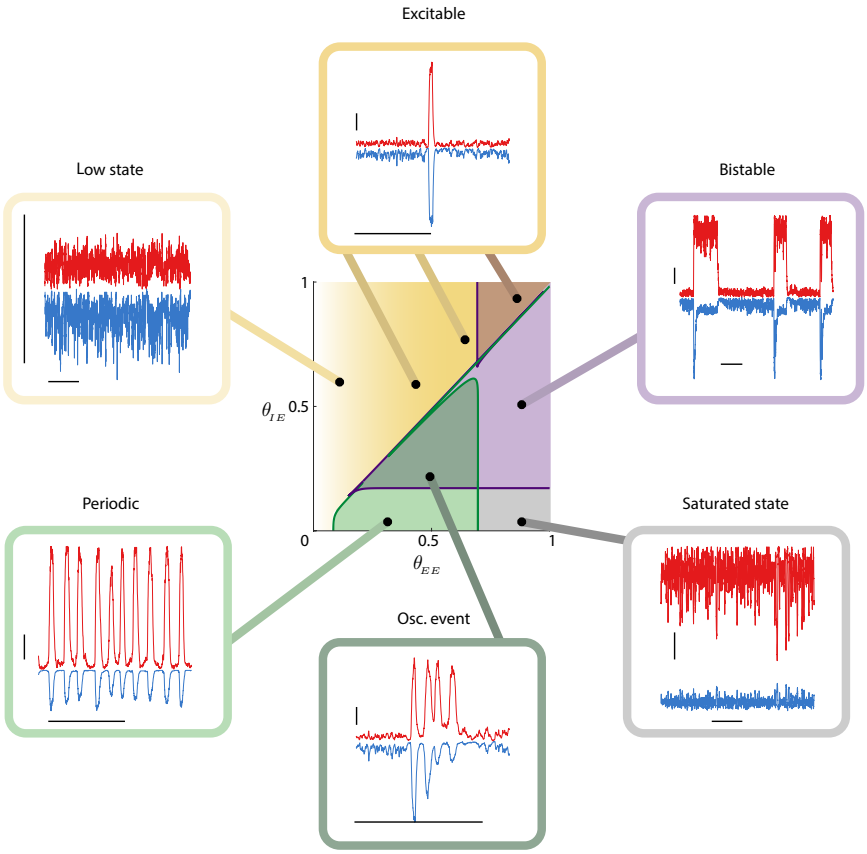

Figure 10: Suite of network dynamics possible in the full model. *Center:* Two parameter bifurcation diagram as the depression thresholds θ_{EE}, θ_{IE} are varied. The purple line marks the location of the saddle node bifurcation (cf. Figure 8) marking the boundary of bistability. To the right of this line (purple shaded regions) the system is bistable. The green line marks the location of the Andronov-Hopf bifurcation (cf. Figure 9) marking the boundary for oscillatory behavior. To the left of the green line but the right of the purple line (green shaded regions), the system can exhibit slow oscillations. Above the unit diagonal, the system is excitable but non-oscillatory (yellow shaded gradient). For low enough values of θ_{EE} the vector field loses its excitatory nature and events do not happen. *Outer ring:* Example simulations showing the suite of possible behaviors. We note that the dark green region (labelled Osc. event) can exhibit *both* oscillatory and non-oscillatory dynamics. Horizontal bars denote 1s, and vertical bars denote 0.2 (arbitrary units). Simulations are for $K = 400$, and, starting with the bistable region and moving clockwise, the threshold values are $(\theta_{EE}, \theta_{IE}) = \{(0.8, 0.5); (0.8, 0.05); (0.5, 0.2); (0.3, 0.05); (0.1, 0.6); (0.8, 0.9)\}$.

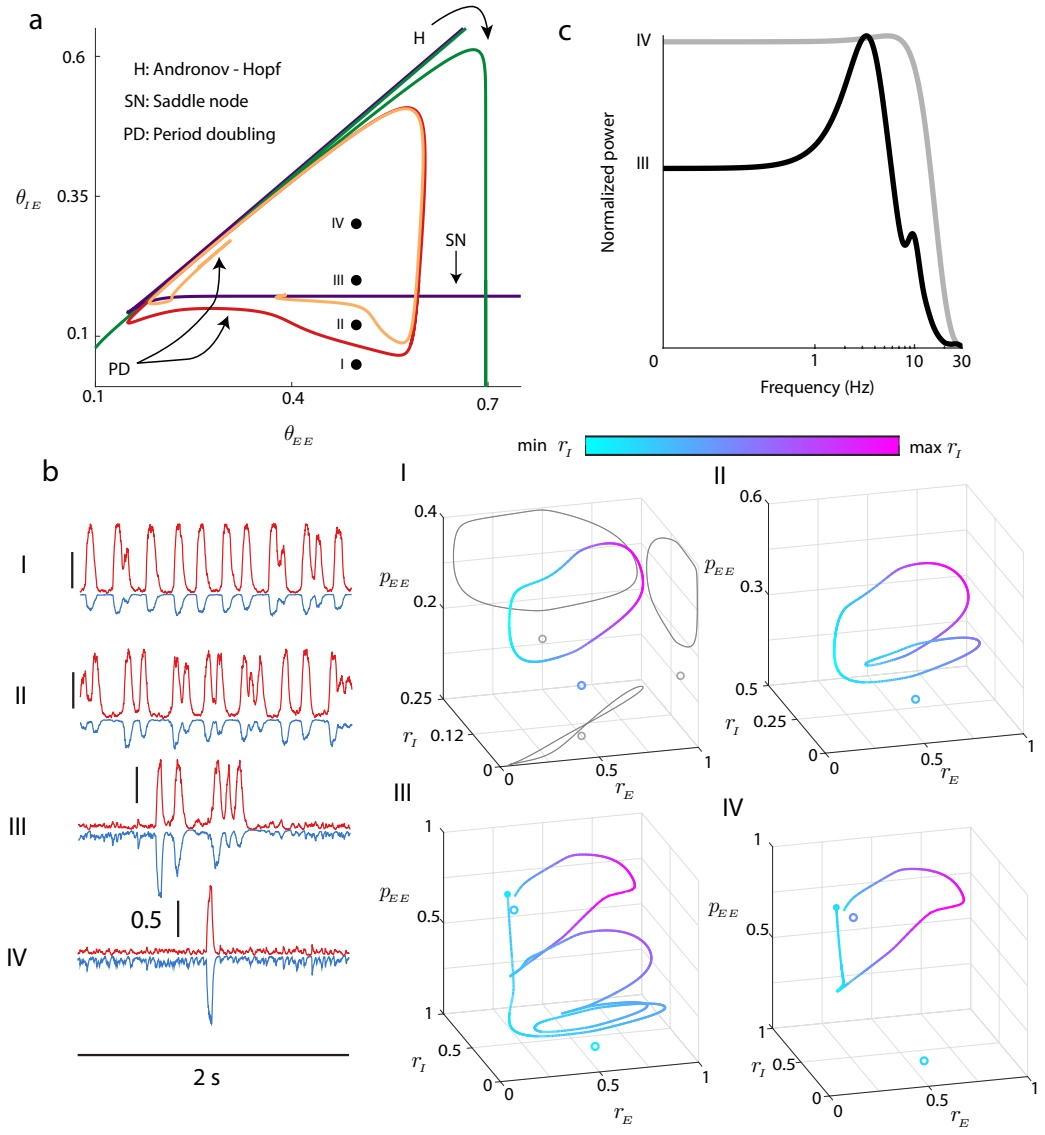

Figure 11: Detailed bifurcation structure. **a:** Magnification of bifurcation diagram shown in Figure 10, with period doubling bifurcations added (red/orange curves). Four different parameter choices are marked ($\theta_{EE} = 0.5, \theta_{IE} \in \{0.05, 0.12, 0.2, 0.3\}$). **b** *Left:* Sample time courses of noisy system for the four parameter choices. *Right:* Trajectories from deterministic system, displayed in (r_E, r_I, p_{EE}). Color denotes depth (location in r_I). Open circles denote unstable fixed points, filled circles denote stable fixed points. For I we show projections onto each of the coordinate planes (gray curves). For choices I and II, we observe a stable limit cycle. For choices III and IV, after the saddle node bifurcation, the only attractor is the stable fixed point, but the echo from the limit cycle allows for complex transients in the excitable system.

2.2.1 Synaptic depression and population events in the auditory cortex

Population recordings from the auditory cortex of both anesthetized and unanesthetized animals show neuronal activity that transitions between desychronized and synchronized states [54, 106–108]. In the synchronized state spontaneous population activity produces periods of brief coordinated spiking activity that is akin to population event behavior [47]. Further, *in vivo* whole cell recording shows sporadic barrages of excitation that are expected from population event dynamics [46], and are similar to those of the *in vitro* data presented in this study (Figure 3). Unsurprisingly, several past modeling studies have thus proposed core mechanisms for population event activity in the auditory cortex.

While our study focuses upon internally generated population event dynamics, the auditory cortex nonetheless robustly produces population events as an onset response to sustained pure tone inputs [109, 110]. It is likely that the physiological mechanisms underlying evoked and spontaneous population events are very related. In tone evoked responses, synaptic depression of excitatory-to-excitatory connectivity has long been thought to play a central role, in part from the well known forward masking effect in auditory cortex [111]. Forward masking is the phenomenon by which cortical spiking responses to successive tone inputs show a strong suppression for later tones, and this suppression lasts for hundreds of milliseconds. *In vivo* whole cell recordings show that the timescale of recurrent inhibition is too fast to explain the suppression phenomenon [109]; however the recovery timescale for excitatory synaptic depression is much longer [89, 112] and matches the suppression recovery time. Thus motivated, previous recurrent circuit models have captured tone evoked population event dynamics strictly through short term synaptic depression of excitatory-to-excitatory connections [84, 85]. In spontaneous conditions these models produce sporadic population events with a single excursion in activity, akin to much of our *in vitro* data set (Figure 4c) and our network model with $\theta_{EE} < \theta_{IE}$ (Figure 10, top inset).

Our *in vitro* recordings also clearly show a subset of population events with rhythmic dynamics. This behavior was not discussed in past experimental studies (including [42]), and to our knowledge represents a novel finding. Our data is consistent with *in vivo* recordings from auditory cortex where population responses show an augmentation of population

responses to successive tones (as opposed to a forward masking) if tone pulsing is restricted to low frequencies [113]. This suggests a network susceptibility for low frequency inputs, of which the spontaneous population event rhythmic dynamic may simply be a reflection. Model networks with strictly excitatory neurons having synaptic depression can show a low frequency ($\sim$ 10 Hz) population oscillation [95, 96]. However, the oscillation in these models is sustained, rather than a transition between a low activity regime and a brief, rhythmic population event. In our model the oscillatory dynamic is mechanistically similar to those of these past studies (Figure 9), however strong recurrent inhibition is needed to create an additional stable low activity state (Figure 8). This complicates how a population event is initiated, since the stable state is not depression stabilized, as was the case in the past models [84, 85, 96].

We use slow synaptic depression of $E \rightarrow I$ synapses to create a saddle point that separates stable low activity from high activity dynamics, and internally generated activity fluctuations prompt stochastic crossings of the stable manifold of the saddle point. For this to be the mechanism for population event initiation our model has a strong requirement: the depression of $E \rightarrow I$ synapses must be recruited at lower excitatory rates than the depression of $E \rightarrow E$ synapses ($\theta_{EE} > \theta_{IE}$ in Figure 10). Fortunately, there is strong evidence for this from paired recordings in layers II/III and IV of auditory cortex (Figure 5a and [89, 90]). There is a precedence of short term plasticity being determined by the post-synaptic target [59, 114], so that a difference between θ_{EE} and θ_{IE} is not curious. However, the θ_{EE} measured in auditory cortex [89] is higher than that generally reported in the neocortex [64, 115], making it easier for $\theta_{EE} > \theta_{IE}$ to be satisfied. This difference in θ_{EE} across datasets may be region dependent, though a likely cause is a shift in short term plasticity over the critical development period [89, 90], making θ_{EE} and θ_{IE} age dependent. Indeed, the recordings we based our measurements of θ_{EE} and θ_{IE} upon were recorded in animals that were postnatal day 19-29 [89, 90], compared to the postnatal day 13-15 used in other studies [64]. In fact, when recordings are restricted to the auditory cortices of younger animals the θ_{EE} and θ_{IE} values are lower [89, 90], consistent with other datasets. In sum, we view the measurement of $\theta_{EE} > \theta_{IE}$ in auditory cortex as a strong postdiction of our model, since the fact that $\theta_{EE} > \theta_{IE}$ was unremarked upon in past studies [89, 90].

2.2.2 Balanced networks and population events

The theory of balanced networks is successful at explaining several core features of cortical activity, namely irregular spiking activity [73], large firing rate heterogeneity [55], and asynchronous population activity [54]. More recently, balanced networks have also been shown to produce stimulus tuned responses in randomly wired networks that lack columnar structure [116] and that they can be critical for selective population codes [117, 118]. Despite these advances, a linear stimulus-response relationship for balanced networks is enforced by the balance condition (see Eq. (2.1)), and this continues to be a large barrier towards this model framework explaining nonlinear properties of cortical response [119].

Mongillo et al.[57] used short term plasticity to impart new nonlinearities to population solutions of balanced networks. In their work they considered facilitation of the $E \to E$ synapse, which produced a network with stable low and high activity states. This is ideal for neuronal dynamics where integration dynamics are required, such as in models of working memory [120]. By contrast, we use depression of the $E \to I$ pathway to give bistable dynamics. In and of itself this is not particularly novel since the removal of recurrent inhibition through a depression-mediated weakening of $E \to I$ pathway is similar, in spirit, to a strengthening of the $E \to E$ pathway via facilitation. However, when $E \to E$ depression is also considered the high state is destabilized and result in an excitable network capable of population event dynamics. This is a population dynamic not previously explored in networks with large recurrent inhibition, instead restricted to networks where inhibition is often ignored [84, 96]. Capturing population events in inhibitory stabilized networks offers an important bridge between cortical recordings where recurrent inhibition is a critical aspect of network function [44, 45, 86–88, 121] with recordings where brief periods of unchecked excitation are clearly apparent [81, 82]. Finally, linking population event initiation to $E \to I$ depression allowed $E \to E$ depression to drive rhythmic dynamics within the spike. Rhythmic population events represent a new population dynamic, not previously discussed in past models, yet it captures a salient feature of our *in vitro* dataset. In total, our work then broadens the repertoire of nonlinear population dynamics that balanced networks can capture.

2.3 METHODS

2.3.1 Markov model of network

For the qualitative characteristics we are interested in studying, it is sufficient to study the aggregate population level behavior of a network rather than the specific microscale instantiation. We model a two population (excitatory, E, or inhibitory, I) network of binary neurons that can be in either an active (1) or inactive (0) state. We let $n_\alpha(t)$ be the number of active neurons in population α at time t, and N_α be the total number of neurons in population α. The firing rate (or mean activity) of each population is calculated as the number of neurons in the active state normalized by the total number of neurons in that population,

$$r_\alpha = \frac{n_\alpha}{N_\alpha} \qquad \alpha \in \{E, I\}. \tag{2.4}$$

Finally, we introduce the probability $P(n, m, t) = \Pr\{n_{_E}(t) = n, n_{_I}(t) = m\}$. We ignore simultaneous transitions and let the probability evolve according to the birth-death process

$$\begin{aligned}\frac{dP(n,m,t)}{dt} &= \omega_+^E(n-1,m)P(n-1,m,t) + \omega_-^E(n+1,m)P(n+1,m,t) \\ &\quad + \omega_+^I(n,m-1)P(n,m-1,t) + \omega_-^I(n,m+1)P(n,m+1,t) \\ &\quad - \left[\omega_-^E(n,m) + \omega_+^E(n,m) + \omega_-^I(n,m) + \omega_+^I(n,m)\right] P(n,m,t)\end{aligned} \tag{2.5}$$

where ω_+^α is the transition rate for n_α increasing and ω_-^α is the transition rate for n_α decreasing. The boundary conditions are chosen as $P(-1, m, t) = P(n, -1, t) = 0$, with forced upper bound $P(N_E + 1, m, t) = P(n, N_I + 1, t) = 0$. The lower boundary condition is straightforward, as it simply indicates that if 0 neurons are in the up state, there are none left to decay downward. The upper boundary condition requires a bit more thought. In the derivation of the mean field equations [103], we formally require N_α to become unbounded. As such, it is technically possible to have a population activity of greater than 1, though this is a non-physical interpretation of the system. By ensuring that the fixed points of the system never get too close to the boundaries, we effectively preclude this possibility, though we include the upper boundary condition as a safeguard. The excitatory transition rates are

taken as

$$\omega_+^E(n,m) = N_E f\left(\sqrt{N_E}\left(j_{EE}p_{EE}\frac{n}{N_E} - j_{EI}\frac{m}{N_I} + I_E\right)\right), \qquad \omega_-^E(n,m) = n \tag{2.6}$$

and the inhibitory transition rates are taken as

$$\omega_+^I(n,m) = \frac{N_I}{\tau_I} f\left(\sqrt{N_I}\left(j_{IE}p_{IE}\frac{n}{N_E} - j_{II}\frac{m}{N_I} + I_I\right)\right), \qquad \omega_-^I(n,m) = \frac{m}{\tau_I} \tag{2.7}$$

where $j_{\alpha\beta}$ is the strength of connection from population β to population α, I_α is directly applied current to population α, p_{EE} and p_{IE} are the plasticity variables, τ_I is the inhibitory timescale, and f is a sigmoidal transfer function. Following the standard procedure [122, 123], this leads to a mean field of the form

$$\begin{aligned} \dot{r}_E &= -r_E + f\left(\sqrt{N_E}\left(j_{EE}p_{EE}r_E - j_{EI}r_I + I_E\right)\right) \\ \tau_I\dot{r}_I &= -r_I + f\left(\sqrt{N_I}\left(j_{IE}p_{IE}r_E - j_{II}r_I + I_I\right)\right) \end{aligned} \tag{2.8}$$

For concreteness, we took the transfer function f to be

$$f(x) = \frac{1}{1+e^{-x}}. \tag{2.9}$$

We note that we have made the assumption that $r_\alpha = n_\alpha/N_\alpha$ as negligibly small (in agreement with the technical requirement that N_α be unbounded), which results in a mean field of the form in Eq. (2.8). If we did not make this assumption, the transfer function term in Eq. (2.8) would have a multiplicative factor of $1 - r_\alpha$ in front of it. We took the excitatory timescale to be unity, and other parameters as $\tau_I = 1$, $j_{EE} = 2$, $j_{EI} = 1$, $j_{IE} = 5$, $j_{II} = 2$, $I_E = -0.12$, $I_I = -0.2$ (see Table 2). To assign a time value to the simulations, we interpret 1 time unit of simulation to be equivalent to 10ms of real time.

Tracking the total number of neurons in the active state for this model is known to reproduce the asynchronous statistics of an Erdős-Renyi network of binary neurons. For the qualitative characteristics we are interested in studying, it is sufficient to study the aggregate quantities n_α and r_α. We also note that for a dense network, the average number of connections K scales with the system size N such that $K/N \to c$, $c \in (0,1)$ a constant. A sparse network has $c = 0$, and a fully connected network has $c = 1$. As such, the "system size" parameter N is easily related to the average number of connections K.

Parameter	Value	Description
τ_E	1	characteristic timescale, excitation
τ_I	1	characteristic timescale, inhibition
j_{EE}	2	connection strength, $E \to E$
j_{EI}	1	connection strength, $I \to E$
j_{IE}	5	connection strength, $E \to I$
j_{II}	2	connection strength, $I \to I$
I_E	-0.12	static input drive to E population
I_I	-0.2	static input drive to I population
K	400	average number of input connections
τ_r	40	characteristic timescale, recovery from depression
τ_d	10	characteristic timescale, recruitment of depression
θ_{EE}	$[0, 1]$	depression threshold, $E \to E$ synapse
θ_{IE}	$[0, 1]$	depression threshold, $E \to I$ synapse
m_{EE}, m_{IE}	2	controls lowest possible value of p_{EE}, p_{IE}
β_{EE}, β_{IE}	50	controls steepness of plasticity curve (cf. Figure 5)

Table 2: Chapter 2 parameters

2.3.2 Model of a synapse

We consider a phenomenological model of a synapse similar to that seen in [64]. Since we are modeling only the mean activity of the network, only the mean synaptic efficacy will affect the governing equations [67]. The dynamics of a synaptic efficacy variable, p, are governed by

$$\frac{dp}{dt} = \frac{1-p}{\tau_r} - \frac{a(r_{_E})p}{\tau_d} \tag{2.10}$$

$$a(r) = \frac{m}{1+e^{-\beta(r-\theta)}} \tag{2.11}$$

The functional form of $a(r)$ was chosen to reproduce the qualitative feature of depression. If we denote the fixed point of Eq. (2.10) as $\hat{p}$, then when $r \sim 0$, $\hat{p} \sim 1$ and as $r \to 1$, $\hat{p}$ approaches a low value (~ 0.1 in practice). The synaptic parameters were chosen as identical for both kinds of synapse, with the exception that the thresholds $\theta_{_{EE}}$, $\theta_{_{IE}}$ were different, and allowed to vary in the case of the bifurcation diagrams. We took $\tau_r = 40$,$\tau_d = 10$,$m = 2$, $\beta = 50$. The threshold parameters $\theta_{_{EE}}, \theta_{_{IE}}$ were bounded between 0 and 1. The ordering of synaptic depression shown in Figure 5 is equivalent to $\theta_{_{EE}} > \theta_{_{IE}}$.

2.3.3 Simulation

The simulations of the noisy system were done by simulating the Markov process governed by Eq. (2.5) to (2.7), where the plasticity variables were treated as slow variables that evolved deterministically between birth/death events (see [124] for some remarks on dealing with hybrid stochastic systems). This is equivalent to simulating a piecewise deterministic Markov process with constant propensities between birth/death events. Since the typical timescale of the plasticity variables is an order of magnitude larger than the typical timescale of the rate variables, this seems a reasonable approximation. Alternatively, a back of the envelope calculation shows that for $\Delta t \sim \mathcal{O}(1/N)$, a typical change in plasticity is

$$\Delta p \sim 1 - \exp(-\alpha \Delta t) \sim \alpha \Delta t \sim \frac{\tau_d + \tau_r}{\tau_d \tau_r} \cdot \frac{1}{N} \sim \mathcal{O}\left(\frac{1}{10N}\right) \sim \mathcal{O}\left(\frac{1}{4000}\right)$$

for our typical parameter values, suggesting that the plasticity variables, and thus propensities, are effectively constant between birth/death events.

For a fixed value of r_E, we can solve Eq. (2.10) for a closed form solution of $p(r_E)$. Then our algorithm for the simulation is as follows:

1. Choose values for n_E, n_I as initial conditions.
2. Initialize r_E, r_I, p_{EE}, and p_{IE}, where $r_\alpha = n_\alpha/N_\alpha$ and $p_{\alpha\beta}$ is chosen as the fixed point of Eq. (2.10) evaluated at r_E.
3. Calculate the transition probabilities $T^E_{+/-}$, $T^I_{+/-}$, and construct the Gillespie probability intervals.
4. Find the time of the next ("current") event.
5. Calculate the current values of p_{EE}, p_{IE} based on the current values of r_E, r_I and Δt, the time between the current event and the previous event.
6. Decide which type of event happened: (in/de)crementing $n_{E/I}$.
7. Update the rate variables r_E, r_I.
8. Return to step 3 and repeat until the simulation is complete.

2.3.4 Bifurcation diagrams

The bifurcation diagrams were computed using XPPAUT [125], simulating the full deterministic system given by Eq. (2.8) to (2.11). A full discussion of computing bifurcations using numerical continuation is a subtle and complicated one beyond the scope of this paper, though XPPAUT is a standard tool for this purpose in applied dynamical systems. Briefly, we highlight the bifurcations of interest to this work. A *saddle node bifurcation* is when a pair of fixed points, one stable and one unstable, annihilate each other (cf Figure 8, 10 and 11). An *Andronov-Hopf bifurcation* is the emergence of a limit cycle when a fixed point changes stability via a pair of purely imaginary eigenvalues. A *period-doubling bifurcation* is when the period a limit cycle doubles, and introduces a characteristic winding of the limit cycle over itself. After such a bifurcation, every other peak returns to the same place is phase space (rather than every peak). In particular, a cascade of period doubling bifurcations over a short parameter region is often associated with the onset of deterministic chaos.

2.3.5 Event detection in data

Please see [42] for a more detailed description of the collection methods. Briefly, we investigate spontaneously active *in vitro* slices from mouse auditory cortex through paired whole cell patch-clamp recordings. Figure 3f was reproduced from [42] by repeating the analysis performed there.

Since recorded membrane potentials and synaptic currents are nonstationary and fluctuate over a wide range of timescales, recorded data was detrended prior to event detection. The detrended data point at time t, $\hat{x}(t)$, was calculated as

$$\hat{x}(t) = x(t) - (\{x(t)\}_{t-w/2,t+w/2} - \langle x \rangle_L) \tag{2.12}$$

where $x(t)$ is the raw data point at time t, $\langle x \rangle_L$ is the average of the entire recording of length L, $\{x(t)\}_{t-w/2,t+w/2}$ is the median of a window of width $w = 3s$ centered at each data point.

To isolate the events, we performed two passes through the data. In the first pass, we isolated potential excursions by looking for periods where the local activity differed from the overall detrended recording average by 5 mV in current-clamp recordings or 0.15 nA in the voltage-clamp recordings. These periods were padded by 200 ms before and 500 ms after the threshold crossing. In the second pass, the specific event times were then marked by hand on an event-by-event basis, where the beginning of the event was taken to be 200 ms before the initial upstroke of the excursion (easily identified by eye), and the end of the event was marked by the last point before the data returned to having a roughly flat slope ($> 1s$ "flat" region required).

Visual inspection of the events show that a nontrivial fraction (approx. 1/3) display oscillatory behavior during the excursion. To compute the power spectra we first denoised the data by applying a Savitsky-Golay filter with a first-order polynomial and a window size of 100 data points (a region approximately 10 ms in width) to remove power over $100Hz$. For each event, MATLAB's detrend function was used to remove any linear trend and center the data, to remove excess low power noise. Finally, the power spectrum of each event was computed using MATLAB's pwelch algorithm, which was then normalized to have unit area.

The power spectra reported for each slice are the average over all events, then normalized to have maximum value 1, for ease of visualization in the heatmaps. The power spectra for the nonevent periods were computed identically, using each nonevent period of length greater than 100ms.

For Figure 4f, there is no natural criteria for choosing the labels "Cell 1" and "Cell 2", since some recordings are for pairs of excitatory currents, some for pairs of inhibitory currents, some recordings have 3 cells simultaneously recorded, etc. As a result, there are a plurality of possible correlation coefficients. To report a single number, we computed the correlation coefficient for $10,000$ possible labelings, where the labels were assigned to each cell randomly. The number reported in Figure 4 ($\rho = 0.67$) is the mean of the distribution of computed correlation coefficients. A representative scatter plot was chosen for panel f.

3.0 ESCAPE PROBLEMS

3.1 INTRODUCTION

In chapter 2 we built a neuronal model that reproduces the complex dynamical behavior of observed data recorded from auditory cortex, while simultaneously retaining the dynamic tension between the excitatory and inhibitory populations. One of the key features of that model and the associated data set is the existence of spontaneous, aperiodic transitions between two qualitative behaviors. This characteristic is not unique to our dataset, but rather, it is pervasive in computational neuroscience. Though the most likely case is that cortical state is a continuum of neuronal activities and not a discrete set of states[126], there is a preponderance of data and modeling work showing that a framework of discrete transitions can serve as a useful starting point to understand cortical activity.

Perhaps the most well known example of this phenomenon is the existence of UP-DOWN states in cortical dynamics. UP-DOWN dynamics consist of (often irregular, aperiodic) alternation between periods of tonic firing (the UP state) and periods of quiescence (DOWN state)[78, 127, 128]. First observed in the context of slow wave sleep[78], UP-DOWN dynamics have since been observed due to anesthesia[78], *in vivo* under thalamic lesioning[129], in awake rodents absent explicit sensory stimulation[128, 130–135], in both rodents and monkeys performing a perceptual task[136, 137], and in a wide range of cortical slices [138–144].

A preponderance of work on UP-DOWN dynamics supports the hypothesis that the intrinsic cause of transition between these two states is the competition between self excitability and mechanisms of cellular fatigue (e.g. synaptic short-term depression, as we used in chapter 2)[141, 145]. Many computational studies have modeled this dynamic, resulting in regular, oscillatory transitions[146–151]. However, the rhythmicity of UP-DOWN

transitions depends heavily on the specific experimental preparation[152–154]. In particular, stimulation of intracortical neurons can reliably induce transitions[43, 155], sometimes by stimulating only a relatively small number of neurons[156]. This sensitivity to stimulation suggests that spontaneously generated transitions are likely due to fluctuations in neuronal activity. Modeling work on fluctuation driven transitions has been done in the spinal cord[157, 158], in networks of excitatory neurons[159, 160], and more recently, in coupled networks of excitatory and inhibitory neurons[161].

Spontaneous transitions in neuronal data are not limited to UP-DOWN dynamics. More broadly, there are a host of contexts in which we observe some sort of switching phenomena between two states in measured data. One such context is that of perceptual bistability. Perceptual bistability is the phenomenon of ambiguous sensory stimuli giving rise to unpredictable sequences of spontaneous perceptual shift. The most commonly studied examples of this are ambiguous visual stimuli[162–164], though there are examples in both auditory[165] and olfactory contexts[166, 167]. Famous examples of ambiguous stimuli include the visual examples of Rubin's vase[162] (often referred to as the "faces or vases" optical illusion), "My Wife and My Mother-in-Law"[162, 168], and the rabbit-duck[169], while the most well known auditory example is the Shepard tone[165].

The mechanism of perceptual bistability is largely unknown[170–173], but there is ample evidence that perceptual transition can be modulated by intracortical state [174–180](e.g. attention), as well as by seemingly noise-induced spontaneous transitions[181–185]. Computational and modeling work on perceptual bistability typically take the form of neural competition models[186–188], as they are a natural candidate for inducing bistability[189–191].

Another context in which spontaneous transitions are found in neuronal data is through neuronal avalanches. A neuronal avalanche is the widespread propagation of spontaneously correlated neuronal activity[192]. The essential feature of a neuronal avalanche is that the event sizes do not display a characteristic scale, and are instead described by a power law[192, 193]. Avalanches have been observed in both slice and cortical culture[192, 194], as well as in awake monkeys[195]. Avalanches appear to be a somewhat generic property of cortical networks, likely arising from the anatomical development of connections[196, 197].

They have been shown to occur when a network is in a state capable of large dynamic range[198], and presumably offer some computational benefit, as they can efficiently store and transmit information[199], and can exhibit incredible stability[194]. Neuronal avalanches are fundamentally different than UP-DOWN states and perceptual bistability in that there is no clear set of stable attractors, and it is unknown if the underlying dynamics should be interpreted in terms of competition between two stable fixed points.

Regardless of their biological differences, each of these examples share a key feature with the data presented in chapter 2, namely that apparently spontaneous, macroscopically measurable qualitative changes in neuronal activity are randomly produced and regularly measured. This phenomenon is widespread throughout computational neuroscience, and biological systems more broadly. While we feel the biological evidence supporting the study of escape problems is well founded, there is an additional issue to which we wish to draw attention. Because even simplistic models of biological behavior can generate spontaneous transitions in this way, understanding and accounting for the generating mechanism is important if one wishes to limit the existence of these types of transitions. To illustrate this point, we consider as an example a small network of pulse coupled linear-integrate-and-fire (LIF) neurons.

An LIF neuron[3] is one of the simplest mathematical models that captures neuron like dynamics. We let $V_j(t)$ represent the membrane potential of the j^{th} neuron at time t. The dynamical equation for V_j is

$$\frac{dV_j}{dt} = -V_j(t) + \sum_{i \neq j} J_{ji}\delta(t - t_i^s) \tag{3.1}$$

$$V_j(t^-) = V_t \Rightarrow V_j(t) = V_r \tag{3.2}$$

where J_{ji} is the strength of the connection from neuron i to neuron j, t_i^s is the time of the s^{th} spike of neuron i, and V_t,V_r are threshold and reset potentials, respectively. If neuron i is excitatory, $J_{ji} > 0$, if neuron i is inhibitory, $J_{ji} < 0$, and if neuron i does not have a synapse connecting to neuron j, $J_{ji} = 0$. In Figure 12 we show two simulations for different network sizes. In both simulations, interplay between excitatory and inhibitory neurons leads to roughly asynchronous behavior. In both cases, the connection strengths have been scaled so

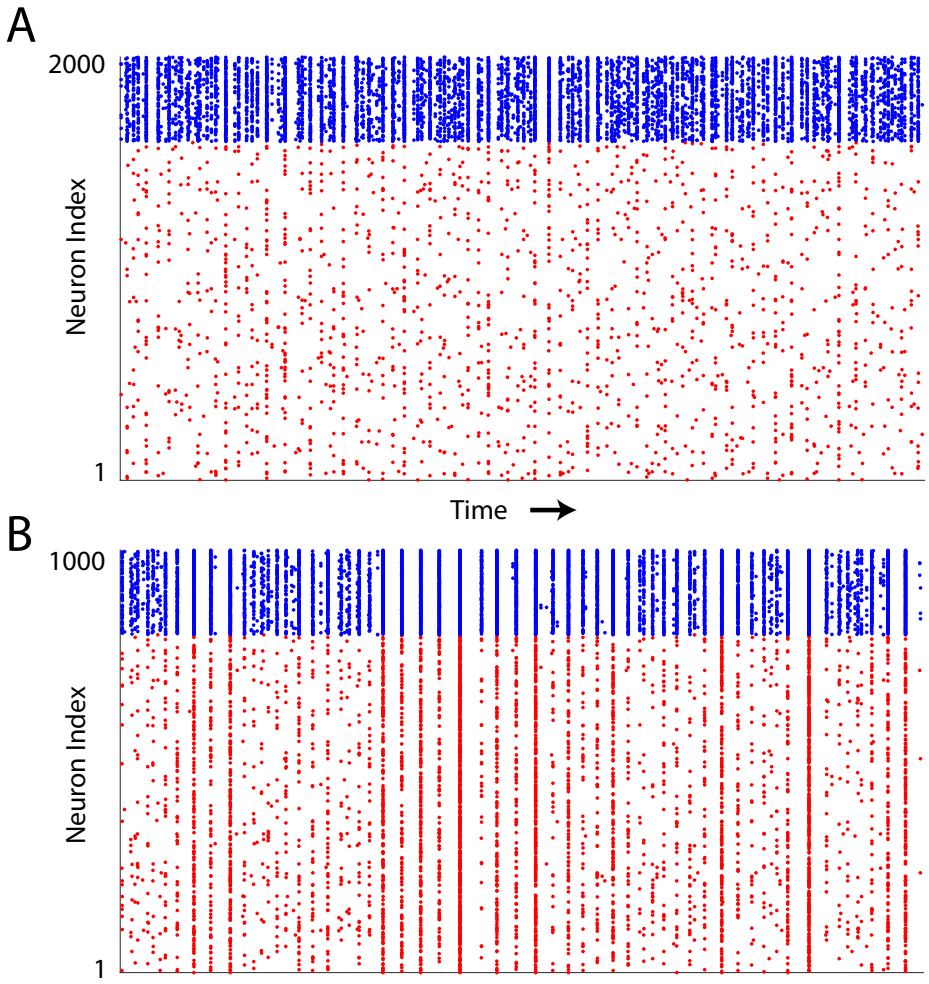

Figure 12: Toy model of spontaneous transitions. Raster plot of two simulations of Erdös-Renyi connected networks of excitatory and inhibitory LIF neurons. Excitatory spikes marked in red, inhibitory spikes marked in blue. Connection strengths are scaled so each neuron in both networks receives the same average input. **A:** $N = 2000$ total neurons. Activity is roughly asynchronous, with sparse bouts of synchrony. **B:** $N = 1000$ total neurons. Activity transitions from synchrony to asynchrony randomly.

that an arbitrary neuron in either network receives the same average input. In Figure 12A, the total system size is $N = 2000$ neurons, and we see that the behavior is roughly asynchronous, with only brief, sparse bouts of synchrony. In Figure 12B, the total system size is $N = 1000$ neurons, and we see clear transitions between qualitatively different dynamic behaviors, switching between regions of synchrony and regions of asynchrony. Because the average input to each neuron is the same in both simulations, individual connections in the smaller network are larger. As a result, individual pulses in the smaller network can have a larger impact on whether or not the efferent neuron spikes, leading to a greater chance for synchronization. This type of effect is what is known as a *finite size effect*, since it will vanish in the $N \to \infty$ limit.

While this model is incredibly simplistic, it highlights the importance of understanding the underlying mechanisms for complex systems. The two simulations in Figure 12 have been controlled to be as similar as possible, with the exception of a single difference, and yet they produce noticeably different behavior. In order to account for these differences, we need to understand the fundamental process that leads to such an issue. This is, in general, an open question. However, the prevalence of such behaviors in biological and modeling systems means that much study has been devoted to this topic, and we have a well-defined framework to discuss finite size effects and noise induced transitions.

3.2 MATHEMATICAL FRAMEWORK

The techniques used in this section, while broadly intuitive, lie outside the standard curriculum of dynamical systems. As such, it is convenient to clearly define some notational conventions and concepts up front, rather than interrupt the thread of the calculation with clarifications for new concepts.

In section 3.1 we gave some evidence for *why* we need a framework to discuss noise in dynamical systems, but the historical process of *how* this framework came about is quite interesting. The earliest studies of noise in physical systems comes from the work of Robert Brown studying the motion of pollen grains suspended in liquid, ca. 1828[200]. It is from

this work that we derive the name *Brownian motion* to describe the most commonly observed and studied stochastic process in the applied sciences. It is also from whence we derive much of the common language to describe stochastic processes of this type, often by referring to the "effective force" on "a particle of negligible mass". One of the most notable advancements in the study of Brownian motion comes from one of Albert Einstein's seminal 1905 papers, where he gave the first rigorous classification of Brownian motion[201]. The formalism typically adopted today to discuss Brownian motion was characterized by Norbert Wiener in the early 1920s[202]. This work has been so instrumental in the development of the study of noise that the standard Brownian motion stochastic process is also know as a *Wiener process*. Interestingly (though not of direct relevance to our work here), Norbert Wiener's personal experiences in World Wars I and II significantly affected the work of applied scientists for years to come. Wiener developed a commonly used signal processing technique (aptly called the *Wiener filter*) while working on the automation of anti-aircraft guns. This work eventually led to his formulating of the field of cybernetics (from the greek word for "steersmanship"), an interdisplinary approach to exploring feedback systems that encourages collaboration between specialists to produce holistic solutions to complicated problems[203, 204].

The two primary intuitive descriptions of a Wiener process are (a) that it is the limit of a random walk[1] or (b) that it is the integral of the standard Gaussian white noise. Historically, the Wiener process has been defined by which Fokker-Planck equation it satisfies, though this definition obscures the most salient features for our analysis. The four conditions listed below are equivalent to the historical definition[205, 206], but more explicitly highlight the features that impact the study of differential equations under noise. For our purposes, we may take a Wiener process $W(t)$ as a stochastic process with:

1. $W(0) = 0$ almost surely
2. independent increments: for $t_0 \geq 0$, $W(t + t_0) - W(t)$ is independent of $W(t_1)$ for any $t_1 < t$.
3. Gaussian increments: $W(t + t_0) - W(t) \sim \mathcal{N}(0, t_0)$
4. continuous (though non-differentiable) paths

[1] Donsker's Theorem, also known as the functional central limit theorem.

Alternatively, we may consider the infinitesimal increment

$$dW(t) = \lim_{\Delta t \to 0} W(t + \Delta t) - W(t), \tag{3.3}$$

and derive the above conditions by taking dW to be delta correlated, i.e. $\langle dW(t_1), dW(t_2) \rangle = \delta(t_1 - t_2)$.

To construct stochastic differential equations, we use a Langevin formulation. This was also originally developed in the context of Brownian motion[207], though here we use slightly more modern notation for clarity. We restrict ourselves to the case of a single spatial dimension for notational simplicity, though the results stated here are easily and straightforwardly extensible to higher dimensions. Similarly, we restrict to the case of first-order differential systems, because higher-order systems can always be reformulated into an equivalent first-order system of higher dimension.

We let $x \in \mathbb{R}$ be a dynamical variable (e.g. firing rate or synaptic efficacy) with dynamics governed by a function $F : \mathbb{R} \to \mathbb{R}$. In the absence of noise, the governing equation is

$$\dot{x}(t) = \frac{dx(t)}{dt} = F(x) \tag{3.4}$$

When white noise is added to this system, the associated Langevin equation may be written as

$$dX(t) = \underbrace{F(X)dt}_{\text{drift}} + \underbrace{\sigma dW(t)}_{\text{diffusion}} \tag{3.5}$$

where σ is the typical magnitude of the noise, and X is a random variable. This is also commonly known as a *drift-diffusion* equation, where the deterministic component F is the drift term, and the stochastic component is the diffusion term.

It is important to draw a subtle distinction between Eq. (3.4) and (3.5). Solutions to Eq. (3.4) may be thought of as trajectories uniquely determined by the initial condition $x(0) = x_0$. In particular, this fact is essential in the numerical study of differential equations, as it ensures that forward integration schemes, with appropriately chosen timesteps, have some hope of converging to the analytic solution. Solutions to Eq. (3.5), on the other hand, are random objects, subject to the influence of $dW(t)$ at every timestep. Solution trajectories are drawn from a distribution, and it is nonsensical to talk of "the" solution. A reasonable

interpretation is to talk of the *distribution* of solutions. With that as a goal, we define the probability density, $\rho(x,t)$ by

$$\rho(x,t)dx = \text{Prob}\left\{X(t) \in (x, x+dx)\right\}. \tag{3.6}$$

A characterization of $\rho(x,t)$ will give a full characterization of $X(t)$.

To investigate $\rho(x,t)$, we need some sort of governing equation to which we can apply differential equations methods. The standard method is to construct a Fokker-Planck equation that governs the behavior of the density. The derivation of this procedure is standard (if somewhat tedious), and can be found in most textbooks on the subject (see, e.g. Risken [205], Gardiner [206]). The end result of this procedure is that for a Langevin equation of the form of Eq. (3.5) the associated Fokker-Planck equation is

$$\frac{\partial \rho}{\partial t}(x,t) = -\frac{\partial}{\partial x}\left[F(x)\rho(x,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[\sigma^2\rho(x,t)\right]. \tag{3.7}$$

Though we have suppressed the notation here, in this derivation we have implicitly assumed that we know the initial condition for x, and have found the equation that governs the evolution of $\rho(x,t)$ *forward* in time. A related question, where we wish to find the evolution of $\rho(x,t)$ *backward* in time given that we know the value of x at a time t, we would find the associated backwards Fokker-Planck equation

$$\frac{\partial \rho}{\partial t}(x,t) = -F(x)\frac{\partial \rho}{\partial x}(x,t) - \frac{1}{2}\sigma^2\frac{\partial^2 \rho}{\partial x^2}(x,t). \tag{3.8}$$

The most substantive difference between these two formations is that the derivatives act only on $\rho(x,t)$ in the backward equation.

Even with a governing equation for $\rho(x,t)$, it is important to note here that often the solution space of the noisy system has no clear analogue to the solutions of the deterministic system. As a simple example, we may consider the deterministic system

$$\dot{x} = -x(x-1)(x+1). \tag{3.9}$$

This system has 2 stable fixed points at $x = \pm 1$ and an unstable fixed point at $x = 0$. The

full solution space of Eq. (3.9) can be classified according to

$$\lim_{t\to\infty} x(t) = \begin{cases} -1 & x(0) < 0 \\ 0 & x(0) = 0 \\ 1 & x(0) > 0 \end{cases} \tag{3.10}$$

Stated another way, the asymptotic average value of $x(t)$ will approach -1,0, or 1 depending on the initial value of x.

In comparison, the stochastic version of this system does not share this property. The stochastic Langevin equation is

$$dX = -X(X-1)(X+1)\,dt + \sigma\,dW \tag{3.11}$$

with associated Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial x}\left[\left(x^3 - x\right)\rho\right] + \frac{\sigma^2}{2}\frac{\partial^2 \rho}{\partial x^2}. \tag{3.12}$$

In this particular case, the stationary solution ρ_s can be found explicitly if we enforce the boundary conditions $\rho_s(+\infty) = \rho_s(-\infty) = 0$ and the normalization condition that ρ_s integrates to 1. In this case, we find

$$\rho_s(x) = C_1 \exp\left(-\frac{x^2(x^2-2)}{2\sigma^2}\right) \tag{3.13}$$

where C_1 is a normalization constant. We see immediately that ρ_s has local maxima at $x = \pm 1$ and a local minimum at $x = 0$. Moreover, since ρ_s is symmetric in x, any odd moment of this distribution will be 0. In particular, this means the average value of X will be 0 for any initial condition, despite the deterministic system having precisely one point which leads to the equivalent condition.

In principle, equations of the form of Eq. (3.5) can be arbitrarily complicated, often through making σ dependent on x and t. This means that in general, even the simplest Fokker-Planck equations can be highly challenging to solve. As an instructive example of the general procedure for dealing with these systems, we now look at a canonical escape problem.

3.2.1 Canonical escape example

Let us first look at a bistable neuronal model to study spontaneous transitions between stable states. Taking the Wilson-Cowan formalism as a starting point, the simplest rate model which exhibits bistability is a single population of excitatory neurons with an external current:

$$\dot{r} = -r + f\left(j \cdot r + I_{app}\right) \tag{3.14}$$

where r is the rate of an excitatory population of neurons, f is a sigmoidal transfer function (in principle the same as f_e and f_i from chapter 2), $j > 0$ is the average connection strength between excitatory neurons, and I_{app} is an applied external current (I_{app} can be positive or negative). To ensure bistability, we need only that $f(x)$ intersect the identity line in three places, which can be ensured through judicious choice of the connection strength j and the applied current I_{app}. Then let us suppose that we are in a regime with two stable fixed points and one unstable fixed point. We will refer to these as the *low rate state*, r_ℓ, the *high rate state*, r_h, and the *saddle*, r_*.

Based on neural data, we expect the rate r to be noisy. From a modeling perspective, we can suppose this noise arises from many possible sources, though throughout this work we will consider it to come from finite size effects. Then our rate equation with noise will be taken as

$$\dot{r} = -r + f\left(j \cdot r + I_{app}\right) + \sigma\xi(t) \tag{3.15}$$

$$= F(r) + \sigma\xi(t) \tag{3.16}$$

where the diffusion coefficient σ scales as $\sigma^2 \sim 1/N$ (with N the system size), and $\xi(t)$ is a white noise term. For notational simplicity, we define $F(r) = -r + f\left(j \cdot r + I_{app}\right)$. Specifically, we note that Eq. (3.16) can be written in of the form of a *Langevin* equation (cf. Eq. (3.5)), as

$$dR = F(R)\,dt + \sigma\,dW \tag{3.17}$$

where R is a random variable taking values in set available to r. For a bistable system, Eq. (3.17) is a special case of a *double well potential*, and we have well-understood methods to study this system. To study transitions between stable states (what is typically called

escape), we first note that for a reasonable function F, we can define an associated potential $\mathcal{U}$ such that

$$F(r) = -\frac{\partial \mathcal{U}}{\partial r} = -\mathcal{U}'(r) \tag{3.18}$$

where $'$ denotes differentiation with respect to the natural argument. This feature of a well-defined potential will not generically be true in higher dimensions, and is the central complication in studying escape problems in multidimensional systems. With such a $\mathcal{U}$, we note that the fixed points r_ℓ, r_h, and r_* are critical points of $\mathcal{U}$, since they are by definition solutions of $F(r) = 0$. Moreover, it is straightforward to show that r_ℓ and r_h will be local minima, and r_* a local maximum (Figure 13A).

Following the general structure of section 3.2, we can define the probability density

$$\rho(r,t)dr = \text{Prob}\left\{R(t) \in (r, r+dr)\right\}. \tag{3.19}$$

which satisfies the Fokker-Planck equation

$$\frac{\partial \rho}{\partial t}(r,t) = \frac{\partial}{\partial r}\left[\mathcal{U}'(r)\rho(r,t)\right] + \frac{\sigma^2}{2}\frac{\partial^2 \rho}{\partial r^2}(r,t). \tag{3.20}$$

We may solve for the stationary density $\rho_s(r)$ by setting $\partial\rho/\partial t = 0$ and integrating directly the corresponding differential equation. We find

$$\rho_s(r) = \mathcal{N} \exp\left(-\frac{2\mathcal{U}(r)}{\sigma^2}\right) \tag{3.21}$$

where $\mathcal{N}$ is found from the normalization condition $\int \rho_s(r)dr = 1$.

In the deterministic version of the question, with no diffusion ($\sigma = 0$), we already know that the minima r_ℓ, r_h are stable, and the saddle point r_* is unstable. Then for any initial condition less than r_*, $r(t) \to r_\ell$, and any initial condition greater than r_* has $r(t) \to r_h$. In the stochastic system, this is no longer the case. For arbitrarily small $\sigma > 0$, the rate is constantly bombarded with noise, and never stays at a single value, so we expect the rate almost never to be near r_*, and to spend most of its time in the bottom of the wells, near r_ℓ, r_h. The smaller σ is, the more true this should be. If we take $\mathcal{U}(r)$ to be sufficiently smooth, we can approximate ρ_s in the regions near the minima. A simple Taylor expansion

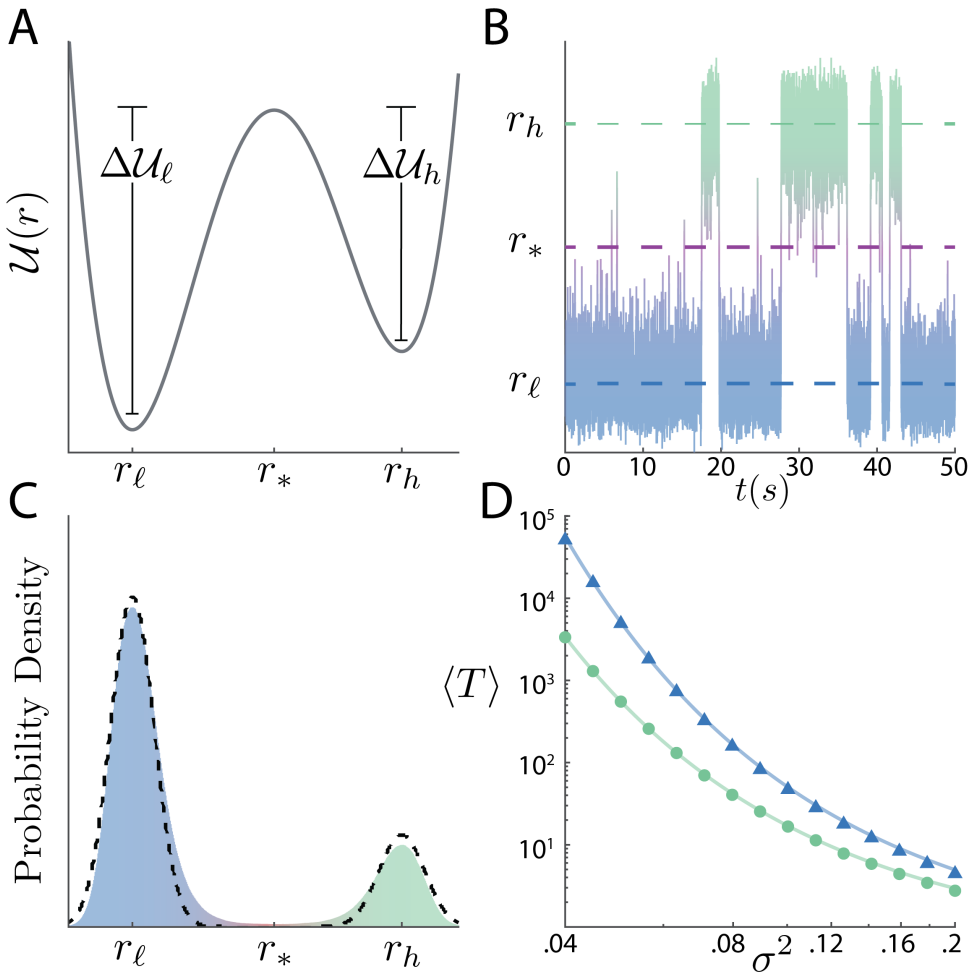

Figure 13: **A:** A potential function with two quadratic wells, depths marked. **B:** Simulation of a rate in the potential from panel A influenced by diffusion. The rate exhibits spontaneous switching between the two stable states. Color of the trajectory indicates which fixed point the rate is near. **C:** A histogram of the long term behavior of the simulation from panel B (blue/green). Dashed line denotes the estimate of the stationary density found from quadratic assumptions on $\mathcal{U}$. **D:** Symbols denote average escape time from r_ℓ (blue triangles) and r_h (green circles), as found by simulations from C. Each value of the diffusion coefficient has a minimum of 1000 trials. Lines denote the Arrhenius estimate.

gives

$$\mathcal{U}(r) \approx \begin{cases} \mathcal{U}(r_\ell) + \frac{1}{2}\mathcal{U}''(r_\ell)(r-r_\ell)^2 & \text{for } |r-r_\ell| \text{ small} \\ \mathcal{U}(r_h) + \frac{1}{2}\mathcal{U}''(r_h)(r-r_h)^2 & \text{for } |r-r_h| \text{ small} \end{cases} \tag{3.22}$$

giving

$$\rho_s(r) \approx \begin{cases} \mathcal{N}\exp\left(\frac{-2\mathcal{U}(r_\ell)-\mathcal{U}''(r_\ell)(r-r_\ell)^2}{\sigma^2}\right) & \text{for } |r-r_\ell| \text{ small} \\ \mathcal{N}\exp\left(\frac{-2\mathcal{U}(r_h)-\mathcal{U}''(r_h)(r-r_h)^2}{\sigma^2}\right) & \text{for } |r-r_h| \text{ small} \\ 0 & \text{otherwise} \end{cases} \tag{3.23}$$

as the stationary distribution. Direct integration gives

$$\frac{1}{\mathcal{N}} \approx \exp\left(\frac{-2\mathcal{U}(r_\ell)}{\sigma^2}\right)\sqrt{\frac{\pi\sigma^2}{\mathcal{U}''(r_\ell)}} + \exp\left(\frac{-2\mathcal{U}(r_h)}{\sigma^2}\right)\sqrt{\frac{\pi\sigma^2}{\mathcal{U}''(r_h)}}. \tag{3.24}$$

If we assume an asymmetry in the depth of the wells (say $\mathcal{U}(r_h) > \mathcal{U}(r_\ell)$, Figure 13A), then for σ small enough one of the terms in $\mathcal{N}$ is overwhelmingly larger than the other, in effect making the deeper well at r_ℓ the more stable of the two states. Correspondingly, we expect the rate to be more likely found in the vicinity of r_ℓ than at r_h (Figure 13C).

With a reasonable characterization of the steady state probability density ρ_s, we are now able to estimate the transition rate between the low and high rate states r_ℓ, r_h. A detailed overview of the background perturbation theory is beyond the scope of this work, but it can be shown[206] that the $t \to \infty$ limit of the small noise perturbation theory does not reproduce the $\sigma \to 0$ limit of the stationary density. This means that for arbitrarily small $\sigma > 0$, an initial condition that starts near r_ℓ will eventually be near r_h. More colloquially, this draws a sharp distinction between deterministic and stochastic systems. In a deterministic system of this form ($\sigma = 0$), initial conditions determine the long time behavior, and an initial condition that starts in the basin of attraction for r_ℓ will never be in the basin of attraction of r_h. In contrast, a metastable system with *any* amount of noise ($\sigma > 0$) will allow the system to (eventually) sample all stable points. We will show that this happens on exponentially long timescales (which is essentially why the linearization of the small noise perturbation theory fails, since $\exp(-K/\sigma^2) \to 0$ faster than any power of σ as $\sigma \to 0$ for fixed K positive).

It will be convenient to work with the conditional probability density $\rho(r,t|y,0)$, speci-

fying that $r(0) = y$ (or, if you like, $\rho(r, 0|y, 0) = \delta(r - y)$). The Fokker-Planck formulation holds for the conditional densities as well. Suppose the rate starts at some $r < r_*$. At a time t, the probability that the rate is still less than r_* is given by

$$G(r, t) \coloneqq \int_{-\infty}^{r_*} \rho(y, t|r, 0)dy. \tag{3.25}$$

If we further suppose that the rate exceeds r_* at time T, then this definition is equivalent to saying $G(r, t) = \text{Prob}\{t \leq T\}$. Moreover, since our system is time inhomogeneous, we have $p(y, t|r, 0) = p(y, 0|r, -t)$. Effectively, rather than thinking of the rate as having a value of r at time 0, we may think of it as having a value of r_* at time T. Then we may use the backward Fokker-Planck equation to see

$$\frac{\partial \rho}{\partial t}(y, t|r, 0) = -\mathcal{U}'(r)\frac{\partial \rho}{\partial r}(y, t|r, 0) + \frac{\sigma^2}{2}\frac{\partial^2 \rho}{\partial r^2}(y, t|r, 0). \tag{3.26}$$

If we then integrate Eq. (3.26) over y, we have the associated governing equation for $G(r, t)$:

$$\frac{\partial G}{\partial t}(r, t) = -\mathcal{U}'(r)\frac{\partial G}{\partial r}(r, t) + \frac{\sigma^2}{2}\frac{\partial^2 G}{\partial r^2}(r, t). \tag{3.27}$$

For the sake of completeness, we give initial and boundary conditions on G. Noting that by definition we have $\rho(y, 0|r, 0) = \delta(r - y)$, the initial condition on G is

$$G(r, 0) = \begin{cases} 1, & -\infty < r \leq r_* \\ 0, & r > r_* \end{cases}, \tag{3.28}$$

or, stated more simply, the probability that $r \in (-\infty, r_*]$ is unity when $t = 0$ (this is a bit of a vacuous statement since the problem was formulated with $r \leq r_*$ at $t = 0$). For the boundary conditions, we need to address each boundary independently. For the right boundary, if $r = r_*$, the rate is 'absorbed' immediately, and so $T = 0$. Thus, $G(r_*, t) = 0$. For the left boundary, it is convention to consider infinite boundaries as reflecting rather than absorbing, since the rate will never leave the interval of interest as it approaches the left boundary. For reflecting boundaries we require a no flux condition, giving us a Neumann

boundary condition, and so

$$\frac{\partial G}{\partial r}(-\infty, t) = \lim_{a \to -\infty} \frac{\partial G}{\partial r}(a, t) = 0.$$

Finally, we can compute the mean exit time by observing the that since $G(r,t) = \text{Prob}\{t \leq T\} = 1 - \text{Prob}\{T < t\}$, we can recover the conditional density by differentiation, and thus we can compute the mean of any function of T by

$$\langle f(T) \rangle = -\int_0^\infty f(t) \frac{\partial G}{\partial t}(r,t) dt \tag{3.29}$$

Then the mean first exit time can be computed explicitly as

$$\begin{aligned} \langle T \rangle &= -\int_0^\infty t \frac{\partial G}{\partial t}(r,t) dt \\ &= \int_0^\infty G(r,t) dt \end{aligned} \tag{3.30}$$

after applying integration by parts. Then integrating Eq. (3.27) over t, we find

$$\begin{aligned} -\mathcal{U}'(r) \frac{\partial \langle T \rangle}{\partial r} + \frac{\sigma^2}{2} \frac{\partial^2 \langle T \rangle}{\partial r^2} &= \int_0^\infty \frac{\partial G}{\partial t}(r,t) dt \\ &= G(r, \infty) - G(r, 0) \\ &= 0 - 1 = -1 \end{aligned} \tag{3.31}$$

with the boundary conditions $\frac{\partial \langle T \rangle}{\partial r}|_{-\infty} = \langle T \rangle|_{r_*} = 0$. Luckily this is a second order linear equation in $\langle T \rangle$, and can be solved explicitly with an integrating factor. Recalling that T is defined as the first exit time for which an initial condition to the left of r_* reaches r_*,we have a fairly simple closed form solution given by

$$\langle T \rangle = \frac{2}{\sigma^2} \int_r^{r_*} ds \exp\left(\frac{2\mathcal{U}(s)}{\sigma^2}\right) \int_{-\infty}^s dz \exp\left(-\frac{2\mathcal{U}(z)}{\sigma^2}\right). \tag{3.32}$$

To get a nicer form for $\langle T \rangle$ we can use information about $\rho_s(r)$ and $\mathcal{U}(r)$. For small σ, $\exp(-2\mathcal{U}(z)/\sigma^2)$ is very small for z near r_*, and we can reasonably approximate the inner integral in Eq. (3.32) by replacing the upper bound s with r_*. This gives

$$\langle T \rangle \approx \frac{2}{\sigma^2} \left[\int_{-\infty}^{r_*} dz \exp\left(-\frac{2\mathcal{U}(z)}{\sigma^2}\right)\right] \int_r^{r_*} ds \exp\left(\frac{2\mathcal{U}(s)}{\sigma^2}\right). \tag{3.33}$$

Moreover, we can further simplify each of these integrals by noting that near the extremal

values $r_\ell, r_*, \mathcal{U}$ is locally quadratic, and the exponential factor will only increase the peakedness of these functions, and away from the extremal values of $\mathcal{U}$, the integrands contribute very little to the integral. Then letting

$$\mathcal{U}(r) \approx \begin{cases} \mathcal{U}(r_\ell) + \frac{1}{2}\mathcal{U}''(r_\ell)(r - r_\ell)^2 & \text{for } r \text{ near } r_\ell \\ \mathcal{U}(r_*) - \frac{1}{2}\left|\mathcal{U}''(r_*)\right|(r - r_*)^2 & \text{for } r \text{ near } r_* \end{cases} \tag{3.34}$$

we can approximate the integrals as

$$\begin{aligned} \int_{-\infty}^{r_*} dz \exp\left(\frac{-2\mathcal{U}(z)}{\sigma^2}\right) &\approx \sqrt{\frac{\pi\sigma^2}{\mathcal{U}''(r_\ell)}} \exp\left(-\frac{2\mathcal{U}(r_\ell)}{\sigma^2}\right) \\ \int_{r}^{r_*} ds \exp\left(\frac{2\mathcal{U}(s)}{\sigma^2}\right) &\approx \sqrt{\frac{\pi\sigma^2}{|\mathcal{U}''(r_*)|}} \exp\left(\frac{2\mathcal{U}(r_*)}{\sigma^2}\right) \end{aligned} \tag{3.35}$$

finally giving the classical Arrhenius formula for escape time

$$\langle T \rangle = \frac{2\pi}{\sqrt{\mathcal{U}''(r_\ell)\,|\mathcal{U}''(r_*)|}} \exp\left(\frac{2\left[\mathcal{U}(r_*) - \mathcal{U}(r_\ell)\right]}{\sigma^2}\right) \tag{3.36}$$

What we have essentially done here is construct a boundary value problem to estimate the total probability density to the left of the saddle (i.e. near the low rate state r_ℓ). The particular methods employed here are especially nice for the canonical one-dimensional example. While the particulars of the calculation need to be adjusted for multidimensional systems, the basic procedure is sound. In particular, there has been much work on formulating and solving the specific boundary value problems to estimate an escape time across a potential barrier for multidimensional, possibly nongradient systems[208–215]. As we will see, there is an additional complication in our model system due to the stochastic hybrid nature of our governing equations.

This process has illuminated the general strategy we wish to take for more complicated problems. Given a deterministic system that exhibits bistability, the noisy analogue of that system will exhibit spontaneous switching between the two stable states (Figure 13B). Moreover, the approximation given by Eq. (3.36) generates quite a nice fit to simulation (Figure 13D), and provides a simple heuristic to estimate escape times. It can be shown that for a large class of stochastic problems, so-called *large deviation principles* can be

derived that give essentially the same qualitative form as Eq. (3.36), suggesting that the intuition used to solve the particle in a double well problem may be applied in a broader context[216]. For an arbitrary system, our procedure is as follows:

1. Given a dynamical system with noise, formulate a governing equation for the probability density $\rho(r,t)$.
2. Using the double-well problem as a guide, construct a potential function, $\mathcal{U}(r)$, that leads to a characterization of the stationary distribution, $\rho_s(r)$.
3. With an initial condition near one of the stable states, estimate the amount of the stationary distribution lying past the saddle r_*, and use this to estimate the probability flux across r_*.
4. Given the probability flux across the saddle, estimate the mean escape time.

For the specific system we wish to study (motivated by our results from chapter 2), there are a number of complications that arise at each step of the proposed procedure.

3.3 NEURONAL MODEL WITH SYNAPTIC PLASTICITY

The model presented in chapter 2 exhibits bistability in certain parameter regimes (cf. Figure 8 and 10). A natural extension of that work is then to investigate the average residence time near the fixed points. In order to attack the simplest problem possible, we make explicit a few assumptions as we reintroduce the model. To retain the balanced character of the model, we require at least two neuronal populations, E and I, as well as strong synaptic weights. Additionally, in order to have multiple stables states, we need at least one synaptic connection to exhibit short term plasticity (Figure 14A).

As before, we consider a phenomenological model consisting of binary neurons. We suppose each population has N neurons, and we track the proportion of active neurons in each population, $r_\alpha(t) = n_\alpha(t)/N$, where $n_\alpha(t)$ is the number of active neurons in population α at time t, $\alpha \in \{E, I\}$. The n_α are Markovian variables. As before, we take the baseline synaptic strength from a neuron in population β to population α to be $J_{\alpha\beta} = j_{\alpha\beta}/\sqrt{K}$,

where $j_{\alpha\beta} \sim \mathcal{O}(1)$ is the unscaled connection strength and K is the average number of connections onto a neuron[48, 54, 55]. Since we are only investigating the aggregate activity, the functional difference between a dense and sparse network is small, so we may take our network to be dense in the sense that K/N is constant. In this way, we may rescale all of the connection strengths $j_{\alpha\beta}$ so as to take $J_{\alpha\beta} = j_{\alpha\beta}/\sqrt{N}$. Finally, we take a sigmoidal activation function f. The specific form of f does not matter for the qualitative features, but in practice we take f to be a logistic function.

To model the effects of short term plasticity, we choose a phenomenological model that acts at the level of rate equations (known to be the limit of a system of individually modeled synapses[67]). We let $p_{IE} \in [0, 1]$ be the synaptic efficacy variable from the excitatory population to the inhibitory population. We choose the dynamics of p_{IE} such that it is a depressing synapse – when firing rates are low, p_{IE} is near 1, and when firing rates are high, p_{IE} is low ($p_{IE} \sim 0.2$ at its lowest value). For large population activity, this system then obeys the following dynamics:

$$\begin{aligned}
\dot{r}_E &= -r_E + f\left(\sqrt{N}\left(j_{EE}r_E - j_{EI}r_I + I_E\right)\right) \\
\tau_I \dot{r}_I &= -r_I + f\left(\sqrt{N}\left(p_{IE}j_{IE}r_E - j_{II}r_I + I_I\right)\right) \\
\frac{dp_{IE}}{dt} &= \frac{1 - p_{IE}}{\tau_r} - \frac{a(r_E)p_{IE}}{\tau_d} \\
a(r_E) &= \frac{m}{1 + e^{-\beta(r_E - \theta)}} \\
f(u) &= \frac{1}{1 + e^{-u}}
\end{aligned} \tag{3.37}$$

Here τ_I is the time constant of the inhibitory population relative to the excitatory, I_α is a source of external drive to population α, τ_r and τ_d are the recovery and depression timescales of the synapse p_{IE}, and θ and β govern where and over how large a region the synapse experiences the full range of depression. For the work that follows, we take $j_{EE} = 2, j_{EI} = 2.4, j_{IE} = 20, j_{II} = 2, I_E = 0.2, I_I = -0.8, \tau_I = 1.1, m = 0.7, \beta = 50, \theta = 0.15, \tau_r = 24, \tau_d = 4$ (see Table 3).

To accurately reflect the stochastic nature of the Markovian n_α, we may define a set of transition rates that have Eq. (3.37) as the expected value of the stochastic process. Given a set of deterministic equations, a set of appropriate transitions rates can be constructed

Parameter	Value	Description
$\tau_{_E}$	1	characteristic timescale, excitation
$\tau_{_I}$	1.1	characteristic timescale, inhibition
j_{EE}	2	connection strength, $E \to E$
j_{EI}	2.4	connection strength, $I \to E$
j_{IE}	20	connection strength, $E \to I$
j_{II}	2	connection strength, $I \to I$
$I_{_E}$	0.2	static input drive to E population
$I_{_I}$	-0.8	static input drive to I population
N	$[10^2, 10^5]$	system size
τ_r	24	characteristic timescale, recovery from depression
τ_d	4	characteristic timescale, recruitment of depression
θ	0.15	depression threshold, $E \to I$ synapse
m	7	controls lowest possible value of p_{IE}
β	50	controls steepness of plasticity curve (cf. Figure 5)

Table 3: Chapter 3 parameters

with some generality [206], but the most natural choice of transitions for this system have the appropriate features [103, 122]. We take as our transition rates

$$\begin{aligned}
(n_E, n_I) \to (n_E - 1, n_I): &\quad \omega_E^- = n_E \\
(n_E, n_I) \to (n_E, n_I - 1): &\quad \omega_I^- = \frac{n_I}{\tau_I} \\
(n_E, n_I) \to (n_E + 1, n_I): &\quad \omega_E^+ = N f\left(\sqrt{N}\left(j_{EE}\frac{n_E}{N} - j_{EI}\frac{n_I}{N} + I_E\right)\right) \\
(n_E, n_I) \to (n_E, n_I + 1): &\quad \omega_I^+ = \frac{N}{\tau_I} f\left(\sqrt{N}\left(p_{IE} j_{IE}\frac{n_E}{N} - j_{II}\frac{n_I}{N} + I_I\right)\right)
\end{aligned} \tag{3.38}$$

In between transitions in the n_α variables, we treat p_{IE} as a deterministic function. This is motivated by the fact that we take the plasticity timescales to be much slower than the excitatory and inhibitory timescales, ensuring that for reasonable N, the rate of individual events is much faster than any other timescale in the system. Systems with these features are sometimes called *stochastic hybrid systems*, or *partially deterministic Markov processes*.

With our choice of parameters, this system displays bistability and spontaneous transitions (Figure 14B). In chapter 2, we made reference to the "low state", in which the rate variables had smaller values than in other regions of phase space. For this chapter, we will primarily be using p_{IE} as the indicator of which state the system inhabits, and when the rates are low the synapse p_{IE} is close to 1. Rather than obfuscate which state is low and which is high, we will refer to the low rate state as the *undepressed* state, and the high rate state as the *depressed* state. Then the bistability of our model system is mainly concerned with three states: the undepressed state, p_ℓ, the depressed state, p_h, and the saddle point, p_*. Because the escape time from the depressed state is exponentially long (make it unobserved in almost all simulations) and requires a multiple order of magnitude increase in simulation time compared to the undepressed state (which itself required hundreds of computation hours to produce Figure 14E). Thus, here we will only investigate the escape time from the undepressed state.

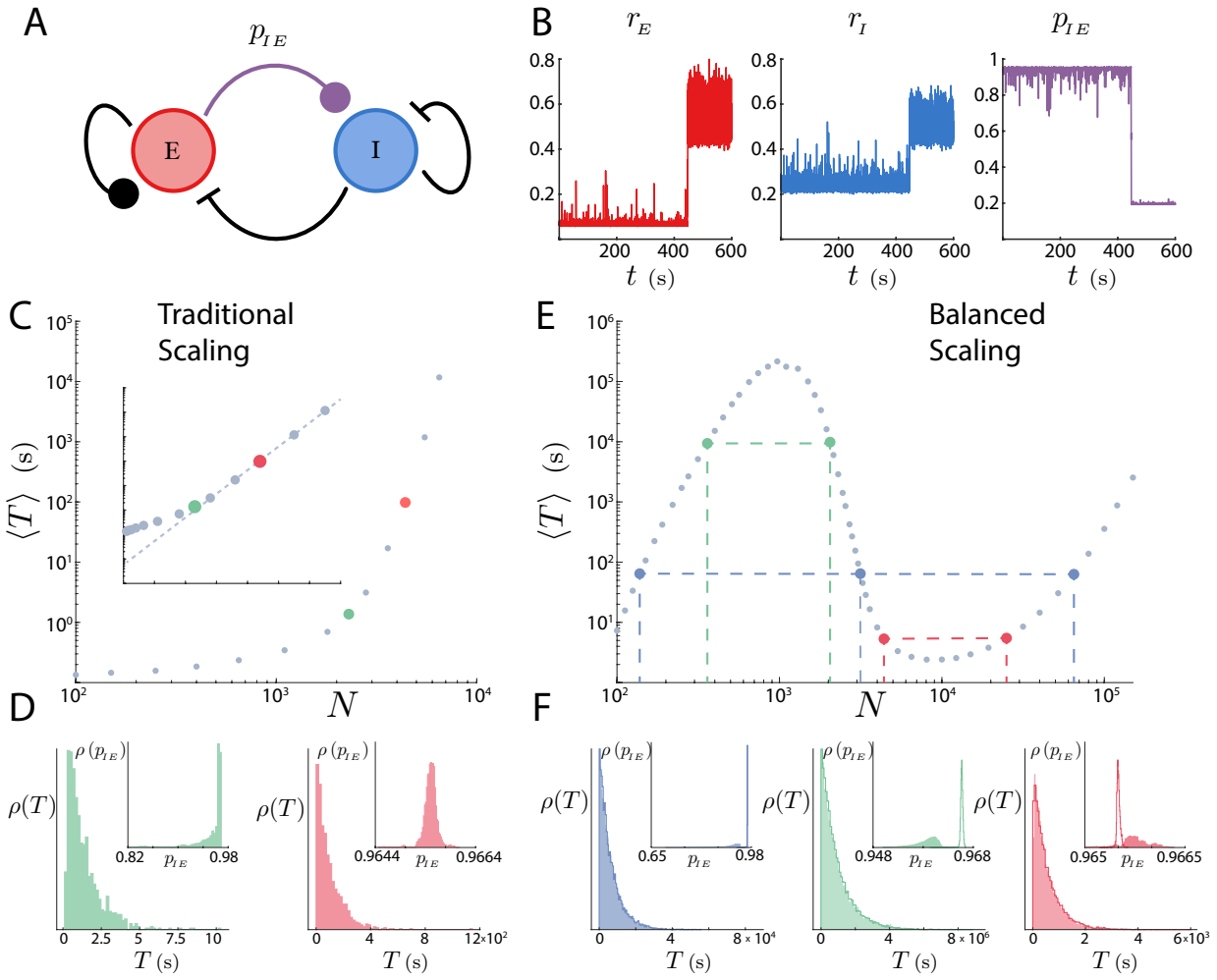

Figure 14: **A:** Schematic of the rate equation formulation, with p_{IE} denoting the plastic synapse. **B:** Simulation of the stochastic system for size $N = 220$ that exhibits bistability and spontaneous transitions. **C:** Simulated escape times, where synaptic strengths are scaled as $1/N$. $\langle T \rangle$ obeys approximately the Arrhenius formula. *Inset:* Plot of same data with only the y axis logarithmic. Dashed line denotes line of best fit for the two largest system sizes, though it passes through almost 6 points. System sizes highlighted are $N = 2300$ (green) and $N = 4400$ (red). **D:** Histograms of exit times for the two points marked in C. *Inset:* Histogram of x-values for the two points marked in C. **E:** The same as C for the balanced scaling case. System sizes highlighted are $N = \{360, 2050\}$ (green), $N = \{4400, 25000\}$ (red), and $N = \{138, 3150, 65000\}$ (blue). **F:** The same as D, for the points marked in E. Larger system sizes are shown as an outline, smaller system sizes are shown as filled.

3.4 COMPARISON BETWEEN SCALING LAWS

Before we construct a governing equation for the probability density of Eq. (3.37), it will first be convenient to compare the behavior of a balanced network with the behavior or a traditionally scaled network. Let us momentarily consider a system analogous to Eq. (3.37) with a traditional Wilson-Cowan style scaling of synaptic strength with system size ($J \sim 1/N$). The deterministic equations for such a system are the same as Eq. (3.37), with the exception that $\sqrt{N}$ does not appear as an explicit parameter anywhere in the equations. We choose synaptic strengths and input currents to be the same as the balanced system when $N = 3600$.[2] The behavior of this system is essentially the same as that of the canonical double well potential example (Figure 14C). To make explicit the parallels, we note that for large N, individual fluctuations in the Markovian variables have a small effect on the rate variables, equivalent to a small diffusion coefficient σ. Similarly, small N corresponds to large σ. Simulations of this system show that $\langle T \rangle \sim \exp(N)$. This is more readily seen by plotting $\log\langle T \rangle$ against N, and observing that as N increases, values of $\langle T \rangle$ converge to a line (Figure 14C inset). Additionally, the probability densities $\rho(T)$ are approximately exponentially distributed, with the variance increasing with system size (Figure 14D). This is the opposite of the trend for the microscale behavior of the state variables. As N increases, fluctuations get smaller and smaller (Figure 14D inset), suggesting a causal relationship between the increase in N and the increase in $\langle T \rangle$. Here we show only the distribution of the synaptic state variable p_{IE}, though the trend is the same for the rate variables. This system behaves essentially as we expect when guided by the intuition of the double well example.

In contrast, something completely different is happening in the balanced scaling case (Figure 14E). The nonmonotonic behavior of $\langle T \rangle$ with N is strikingly and fundamentally different than the intuition from the double well problem. If we select any set of system sizes for which $\langle T \rangle$ is approximately the same, we see that the densities $\rho(T)$ are still exponentially distributed, suggesting that transitions are in fact fluctuation driven events. Moreover, the exponential distributions are N-dependent only so far as $\langle T \rangle$ depends on N. This is shown in Figure 14F, where we plot the distribution $\rho(T)$ for different N, and observe that the

[2] Take $j_{\alpha\beta}, I_\alpha$, and map them to $60 j_{\alpha\beta}, 60 I_\alpha$.

histograms are visually near-identical. However, we note that the microscale behavior varies with system size in the expected way: as N increases, the distribution of observed p_{IE} values $\rho(p_{IE})$ becomes more sharply peaked. Thus there must be some compensatory mechanism in the dynamics of this system that counteracts the decrease in fluctuations to increase the likelihood of an escape event.

At first glance, there are potentially two reasonable guesses for what may be causing this non-monotonicity. Our system differs from the canonical double well example by the fact that it has a partially deterministic component and also that the system size N is an explicit parameter of the governing equations. However, simulations show that the partially deterministic process with traditional scaling laws effectively behaves as a traditional double well (Figure 14C,D), suggesting that the partially deterministic component alone is not enough to cause nonmonotonicity in the escape times.

Our hypothesis is that the fundamental driver of the nonmonotonic behavior is the explicit system size dependence in the governing equations. In this case, the system size actually changes the vector field of the underlying deterministic system, effectively changing the potential landscape of phase space. However, since N is also the source of noise in the system, this system has an intrinsic link between the traditional "drift" and "diffusion" components of Langevin style stochastic processes. To construct a similar condition in the canonical double well problem, we would need to make the well depths $\Delta\mathcal{U}$ functions of the diffusion σ. To the best of our knowledge, this particular problem has not been studied, however, this appears to be a promising direction in understanding the intricate links between finite size driven noise and aggregate dynamical behavior.

3.5 STOCHASTIC HYBRID SYSTEMS AND THE PSEUDOPOTENTIAL

The first hurdle in estimating the escape time for our system is that there is no obvious candidate for a potential function $\mathcal{U}$ that governs the dynamics of this system. In order to address this, it will be beneficial to reformulate our problem slightly. Rather than considering the rate variable as a limit of the Markovian process for large system size (as we effectively

do when we think of an underlying multidimensional dynamical system acting under the influence of noise), we make explicit the discrete nature of our system.

Our state variables $(n_{_E}, n_{_I}, p_{_{IE}})$ live in $\mathbb{S}\times\mathbb{S}\times\mathbb{R}$, where $\mathbb{S} = [0, 1, \ldots, N]$ is a discrete state space. Through its dependence on $n_{_E}$, the synaptic variable $p_{_{IE}}$ defines a related stochastic process $P(t)$, with governing equation given by

$$\frac{dP(t)}{dt} = \frac{1-P}{\tau_r} - \frac{a(n_{_E})P}{\tau_d} \tag{3.39}$$

We may define a joint probability density for our state vector by

$$\rho(p, n, m, t)dp = \text{Prob}\left\{P(t) \in (p, p+dp), n_{_E}(t) = n, n_{_I}(t) = m\right\} \tag{3.40}$$

In section 3.2, we reviewed the Fokker-Planck equation equation which governs the probability density for a Langevin equation. For a stochastic hybrid dynamical system, we begin instead with a differential Chapman-Kolmogorov (CK) equation, which allows us to measure the probability flux through both the discrete and continuous changes in the system.

To compute the probability flow between states due to transitions in the discrete variables, we let A be the transition matrix for the Markov variables $n_{_E}, n_{_I}$ (schematic in Figure 15A). The off diagonal entries of A are filled with the transition rates defined by Eq. (3.38), and the diagonal entries of A are constrained to have column sum 0. By construction, A is a stochastic matrix which defines how n_α move about. There are N^2 possible states for the ordered pair $(n_{_E}, n_{_I})$, which makes A of size $N^2 \times N^2$. Moreover, A is a function of $p_{_{IE}}$, inherited through the transition rates ω.

To account for probability flow due to changes in the continuous variable, we define a diagonal matrix B where the entries are defined according to

$$B_{(n,m)}(p) = \frac{1-p}{\tau_r} - \frac{a(n)p}{\tau_d} \tag{3.41}$$

where (n, m) denotes a single state in the 2 dimensional Markov process, uniquely specifying a row in the matrix B. This ensures B is also of size $N^2 \times N^2$. We note that the right hand side of Eq. (3.41) does not depend on m. This is because we are modeling a plastic synapse from $E \to I$, and expect that the amount of depression should depend on the pre-synaptic rate (cf. Eq. (1.15)). This is consistent with the modeling choices made to integrate short

term plasticity into a mean-field network structure [67].

Rather than writing out a tedious sum over all possible states, we can construct a vectorized form of the CK equation. We define $\Psi(p,t) = (\ldots, \rho(p,n,m,t), \ldots)^T$ to be the vector of densities for each state of the Markov process. Then the differential CK equation is essentially a *master equation*, given by tracking the flow of probability density between states. We have

$$\frac{\partial \Psi}{\partial t} = A\Psi - \frac{\partial}{\partial p}(B\Psi) \tag{3.42}$$

We note explicitly that both A and B depend on p. Eq. (3.42) can be understood intuitively in terms of processes we've already established. The $A\Psi$ term tracks probability flow due to changes in the discrete variables $n_{_E}, n_{_I}$, entirely analogous to a simple discrete Markov process. The second term, involving $B\Psi$, tracks probability flow due to changes in the continuous variable $p_{_{IE}}$. Note specifically that this term shares the general form of the drift term in a Fokker-Planck equation (and is derived through the same procedure). Because the continuous variable $p_{_{IE}}$ has no intrinsic noise (and inherits stochasticity through interacting with the discrete Markovian variables), there is no analogous diffusion term. Equations of this form have been studied reasonably completely [216–219], and there is a well-defined (though time consuming) algorithm to obtain estimates for the mean exit time. Before we compute the mean exit time explicitly, it is helpful to first consider a simpler calculation.

3.5.1 Quasi-steady state approximation

The simplest plausible technique that we could try is a quasi-steady state approximation. There are two reasons to expect that this might be a reasonable simplification. First, the timescales of plasticity τ_r, τ_d are approximately an order of magnitude larger than the timescales of excitation and inhibition (1 and τ_I, respectively), suggesting a natural separation in the effective dynamical timescales of the state variables. Second, and more importantly, an effect of that timescale separation is that the rate of individual transitions of n_α is much faster than any change in the deterministic variable $p_{_{IE}}$, where the typical time between stochastic transitions ($\Delta t \sim \mathcal{O}(1/N)$) induces a change in $p_{_{IE}}$ of $\Delta p_{_{IE}} \sim \mathcal{O}(1/(10N))$. Taking the transitions in n_α as fast, the quasi-steady state approximation is to let $n_{_E}$, $n_{_I}$

relax to their stationary distributions, namely by taking

$$A\Psi = 0 \tag{3.43}$$

for fixed values of p_{IE}. We denote the solution to Eq. (3.43) as Ψ_s. In Figure 15B, we show examples of the stationary distribution Ψ_s for representative values of p_{IE}. Along the top row, shown as a heat map normalized to have area 1, are three instances of Ψ_s. As expected, as p_{IE} decreases, the steady state of the rate variables changes, and as we enter the oscillatory regime for very low p_{IE} the distribution spreads. Along the bottom row, we show the marginal density for n_E, the only discrete variable that directly affects the vector field induced by B. In the undepressed state, we can see that the marginal density $\rho(n_E)$ is approximately Gaussian, but displays decidedly more peakedness in the depressed ($p_{IE} \sim 0.2$) state.

By varying p_{IE} and solving Eq. (3.43), we can define $\Psi_s(p)$, the effective steady state distribution of the discrete variables as a function of the continuous variable. Using $\Psi_s(p)$, we can construct an average vector field governing the dynamics of p_{IE}. To do this, we let $\mathbf{b}$ be the $1 \times N^2$ vector constructed from the diagonal matrix B, and we construct the average vector field by performing a weighted sum over $\Psi_s(p)$ (Figure 15C, for a single value of N):

$$\overline{B}(p) = \mathbf{b}^T \Psi_s(p) \tag{3.44}$$

Through this simplification, we have reduced our system to an effective one dimensional system, with the dynamical variable not being the *rate* (as in Section 3.2.1), but rather the *plasticity state*. The new governing equation for p_{IE} is

$$\frac{dp_{IE}}{dt} = \overline{B}(p_{IE}) \tag{3.45}$$

For this averaged vector field, we define a pseudo-potential $\hat{\mathcal{U}}$ by integrating, i.e.

$$\hat{\mathcal{U}}(p) = -\int_{p_\ell}^{p} \overline{B}(s)\, ds \tag{3.46}$$

where p_ℓ denotes the value of p_{IE} in the undepressed, low rate state, and p denotes the value of p_{IE} in an arbitrary state. Three representative examples of $\hat{\mathcal{U}}$ for three different values of

N are shown in Figure 15D. Following the intuition from section 3.2.1, we expect that the average escape time should be directly related to the well depth of the pseudo-potential $\hat{\mathcal{U}}$. For our system, with the standard parameters (Table 3), this well depth is non-monotonic with the system size N (Figure 15E).

This crude analysis captures the broad qualitative features of the observed escape times (Figure 14E), and provides compelling evidence that our intuition is essentially correct. That is:

1. There is a (pseudo)potential structure to the system that serves as the largest determining factor in setting the average transition time between states.
2. The transitions are reliably noise driven, in that the distribution of escape times is exponentially distributed (even though the variation in the state variables is roughly Gaussian) (cf. Figure 14D,F).
3. Transitions between states can be effectively marked by the passing of the unstable manifold of a saddle point, and
4. the escape time can be estimated by estimating the amount of the stationary distribution that crosses that unstable manifold.

At this point, we can apply the basic methods from section 3.2.1, along with an appropriate estimation of the magnitude of the noise, to estimate the average escape time. However, it is known that applying the Arrhenius estimate with the pseudo-potential $\hat{\mathcal{U}}$ will give inaccurate results for the escape time in a stochastic hybrid system[218]. The discrepancy arises from deviations in the exponential factor of the Arrhenius estimate when the rate is approaching, but still far from, the saddle.

The work of Bressloff & Faugeras [216], Keener & Newby [217], Newby & Keener [218], details the proper approach for quantitative estimates through the use of the Wentzell-Kramers-Brillouin (WKB) approximation in solving an appropriate boundary value problem. We apply a straightforward adaptation of this technique to estimate the escape time for our model system. The main complication in our adaptation comes from the computational complexity of solving for the stationary distribution $\Psi_s(p)$ and estimating an eigenvalue and eigenvector for an exceptionally large matrix.

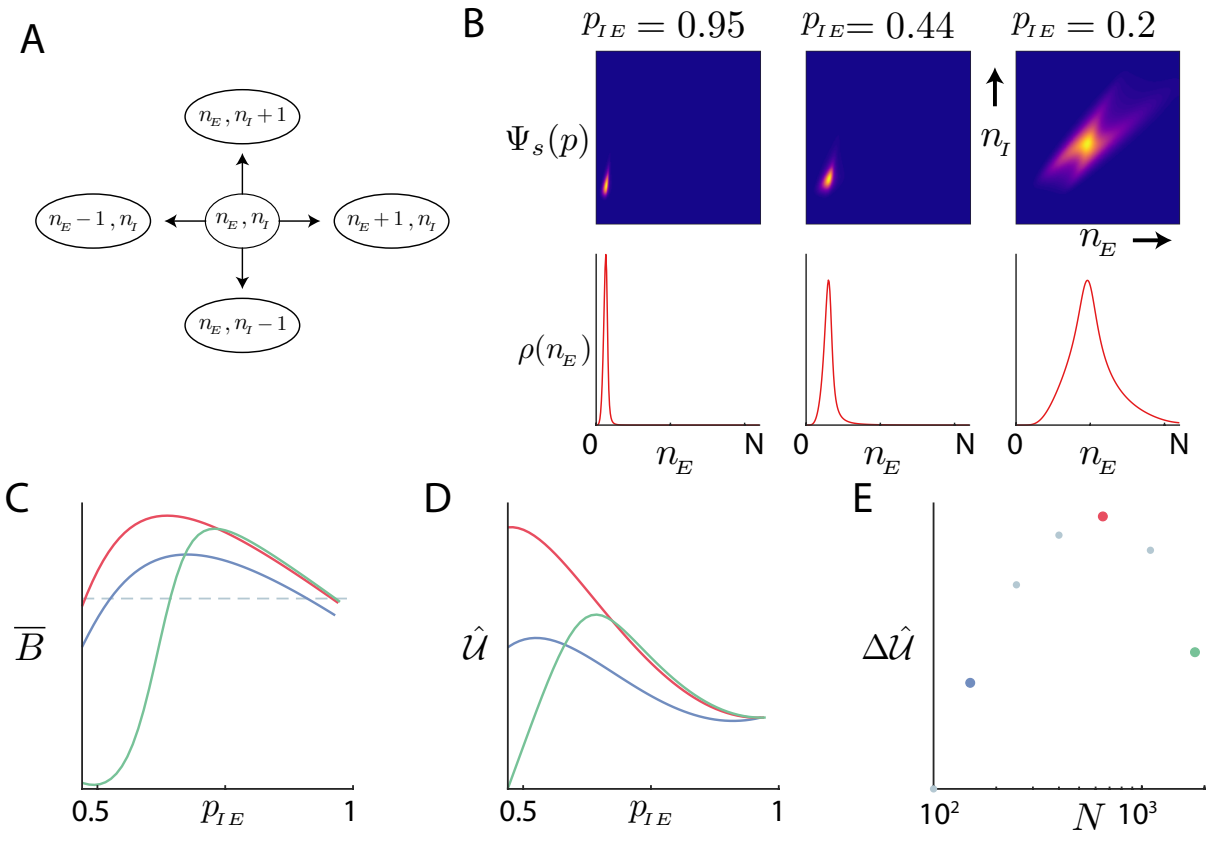

Figure 15: **A:** Schematic of our two dimensional Markov process. **B:** stationary (top) and marginal (bottom) distributions for selected values of x. **C:** Example estimates of the average vector field $\overline{B}$ for 3 different system sizes. Blue represents the lowest system size, red is an intermediate system size, and green is the largest system size. **D:** Example potential functions from the undepressed state to the saddle for 3 different system sizes. Coloring is as in panel C. **E:** Well depths as a function of system size. The potentials found in D are highlighted with matching colors. We observe an internal extremal value.

3.5.2 Escape time estimation

The work in this section follows closely the work developed in Newby & Keener [218]. Rather than simply repeating the calculations performed there (since they were done in great generality), we provide a general overview of the method as it relates to our model system, and highlight the challenges specific to this system.

The first step in estimating the escape time is to set up the appropriate boundary value problem. The construction of such a boundary value problem for escape problems often follows the same general structure [206, 218]. We let T be the random variable equal to the first time at which the plasticity variable p_{IE} reaches the saddle p_*. We wish to approximate the probability distribution for T. In section 3.2.1, we were able to explicitly formulate a boundary value problem for the first moment of this distribution, $\langle T \rangle$. This will not generally be possible, so we instead formulate first a boundary value problem for the full distribution, and then proceed to estimate the first moment.

Suppose the plasticity variable has an initial condition in the undepressed state, $p_{IE} \in (p_*, \infty)$. We define the *survival probability*, $\mathcal{S}(t)$, to be the probability that the plasticity variable p_{IE} is still inside the interval (p_*, ∞) at time t, given by

$$\mathcal{S}(t) \equiv \sum_{(n,m)} \int_{p_*}^{\infty} \rho(p, n, m, t)\, dp. \tag{3.47}$$

We notice that the distribution for T is functionally related to $\mathcal{S}$ by

$$\text{Prob}\{t > T\} = 1 - \text{Prob}\{t \leq T\} = 1 - \mathcal{S}(t). \tag{3.48}$$

To find the density function for T, then, we need only differentiate this distribution, giving $\mathcal{F}(t) = -\partial \mathcal{S}/\partial t$. Then our goal becomes to find an expression for $\partial \mathcal{S}/\partial t$.

We can relate $\mathcal{S}$ to the stationary density Ψ in the following way. By first integrating an arbitrary row of Eq. (3.42) over the interval (p_*, ∞) and then summing over all possible discrete states (n, m), the first term in Eq. (3.42) becomes the time derivative of the survival probability $\mathcal{S}$:

$$\frac{\partial \mathcal{S}}{\partial t} = \sum_{(n,m)} \int_{p_*}^{\infty} \frac{\partial \rho}{\partial t}(p, n, m, t)\, dp \tag{3.49}$$

Similarly, the fact that A is a stochastic matrix (with column sum zero), gives that

$$\sum_{(n,m)} \sum_{(i,j)} \int_{p_*}^{\infty} A_{(n,m);(i,j)} \rho(p,i,j,t)\, dp = 0 \tag{3.50}$$

Finally, applying the boundary condition that $\lim_{p\to\infty} \rho = 0$, the last term from Eq. (3.42) results in

$$\sum_{(n,m)} \int_{p_*}^{\infty} \frac{\partial}{\partial p} B_{(n,m)} \rho(p,n,m,t)\, dp = -\sum_{(n,m)} B_{(n,m)} \rho(p_*,n,m,t) \tag{3.51}$$

Combining each of these results gives us a governing equation for the survival probability. After rewriting summation as vector multiplication, we have

$$\frac{\partial \mathcal{S}}{\partial t} = \mathbf{b}^T \Psi(p_*) \tag{3.52}$$

We have thus reduced the problem of estimating the escape time to one of estimating the distribution Ψ at the saddle. If we can approximate $\Psi(p)$, we can approximate $\partial\mathcal{S}/\partial t$, and thus the density function $\mathcal{F}$, which characterizes the escape time T. The details of this approximation are worked out in great detail in Newby & Keener [218] and Keener & Newby [217], and here we simply apply them to our specific system. Thus, in what follows, we present the main points necessary to apply this method to our system. We refer the interested reader to Newby & Keener [218] for the specific proofs.

The most important point to estimating Ψ is understanding that a stationary solution to Eq. (3.42) will have zero net flux across the saddle point p_*, and so provides an inaccurate estimate of the escape time. However, the natural timescale separation in the system (since the transition matrix A has terms that are $\mathcal{O}(N)$) provides an avenue of attack. The use of spectral projection methods[220] for singularly perturbed boundary value problems, along with the Wentzell-Kramers-Brillouin (WKB) method[221], provide sound estimates for the escape time.

The first key component is to assume that $\Psi(p,t)$ has an eigenfunction expansion of the form

$$\Psi(p,t) = \sum_{j=0}^{\infty} c_j e^{-\lambda_j t} \phi_j(p). \tag{3.53}$$

Because of the separation of timescales, we will have that the principal eigenvalue λ_0 will be

exponentially small compared to all other eigenvalues[217, 222]. We will refer to λ_0 as the *escape eigenvalue*. This means that all higher modes of the eigenfunction expansion can be ignored in the long-time limit, and we will have

$$\Psi(p,t) \sim \phi_0(p) e^{-\lambda_0 t}. \tag{3.54}$$

To find an expression for the exit time distribution, we first note that each eigenfunction in Eq. (3.53) satisfies Eq. (3.42). In particular, Eq. (3.54) then implies that we have the condition

$$\lambda_0 \phi_0 = A\phi_0 - \frac{\partial}{\partial p}\left(B\phi_0\right). \tag{3.55}$$

As we showed in Eq. (3.49) to (3.51), we can transform the right hand side of the above equation into the right hand side of Eq. (3.52). Applying the same procedure to the left hand side, we have

$$\lambda_0 \sum_{(n,m)} \int_{p_*}^{\infty} \phi_0(p)\, dp = \lambda_0, \tag{3.56}$$

where we have used the fact that the probability distribution must normalize to 1. In the long time time limit, our approximation is that $\phi_0(p)$ encompasses all possible states of p_{IE}, and by assumption we are taking the initial condition as somewhere in the basin of attraction of the undepressed state (i.e. above the saddle point p_*).

Then from Eq. (3.52) it follows that the exit time distribution will scale as

$$\mathcal{F}(t) \sim \lambda_0 e^{-\lambda_0 t}, \tag{3.57}$$

which has a first moment of

$$\langle T \rangle \sim \frac{1}{\lambda_0}. \tag{3.58}$$

We have now reduced this problem to one of estimating the escape eigenvalue.

Before we work through this calculation, however, it will first be convenient to understand some specific results from Newby & Keener [218] that apply to our system. As we will see below, the problem of estimating the escape eigenvalue can be framed in terms of a generalized eigenvalue problem. For our system, for almost all values of p, the diagonal

matrix B is invertible. In this case the generalized eigenvalue equation

$$A\mathbf{v} = \mu B\mathbf{v} \tag{3.59}$$

is equivalent to the standard eigenvalue equation

$$B^{-1}A\mathbf{v} = \mu\mathbf{v}. \tag{3.60}$$

It is simple to see that $B^{-1}A$ and A share the same null space (since B is invertible), and that the nullspace is spanned by a unique vector φ_0 (since A is a stochastic matrix – this follows from the Frobenius-Perron theorem). Going forward, we let φ denote *right* null vectors of $A - \mu B$, and η denote *left* null vectors of $A - \mu B$. In particular, we note that φ_0 is a probability vector, meaning that it has sum 1, and all entries are positive.

It can be shown for matrices of the form $B^{-1}A$, with A a stochastic matrix and B defined as in Eq. (3.41)[218] that there is precisely one other positive right eigenvector, which we denote φ_1. The corresponding eigenvalue μ_1 is nonzero for $p \neq p_*, p_\ell$, and has the property $\mu_1(p) > 0$ for $p > p_\ell$ and $\mu_1(p) < 0$ for $p_* < p < p_\ell$. The eigenvectors φ_0, η_0 (with eigenvalue 0) and φ_1, η_1 (with eigenvalue μ_1) are essential to defining an effective potential $\mathcal{U}$ for our system. See Section 3.5.3 for a discussion of the complications of estimating these eigenvalue/eigenvector pairs.

The second key component in estimating the escape time comes from the WKB approximation, which provides a framework to reasonably define a potential function $\mathcal{U}$. (Recall that this is the central difficulty in applying the standard method from Section 3.2.1.) From Eq. (3.54) we see that it will suffice to approximate ϕ_0. We denote this approximation ϕ_ε. Motivated by the intuition from Section 3.2.1, we apply the WKB ansatz, namely that there exists a potential $\mathcal{U}$ such that the ϕ_ε can be expressed as

$$\phi_\varepsilon(p) = \mathbf{g}(p)e^{-N\mathcal{U}(p)} \tag{3.61}$$

where $\mathcal{U}(p)$ is a scalar function and $\mathbf{g}(p)$ is a vector. To construct such an approximation, we take the asymptotic expansions $\mathcal{U} \sim \mathcal{U}_0 + N^{-1}\mathcal{U}_1$ and $\mathbf{g} \sim \mathbf{g}_0 + N^{-1}\mathbf{g}_1$, and substitute these in to Eq. (3.42) (equivalently, Eq. (3.55)). After applying the standard techniques of

perturbation theory, to lowest order we arrive at the matrix equation

$$A\mathbf{g_0} = -\mathcal{U}_0' B\mathbf{g_0}, \tag{3.62}$$

which, for a fixed value of p, is equivalent to the matrix equation Eq. (3.59).

Newby & Keener [218] show that we can approximate this dominant term of the effective potential as

$$\mathcal{U}_0(p) = -\int_{p_*}^{p} \mu_1(y)\, dy \tag{3.63}$$

and show that the first order term of the potential satisfies

$$\mathcal{U}_1'(p) = \frac{\eta_1^T (B\varphi_1)'}{\eta_1^T B\varphi_1} \tag{3.64}$$

where we recall η_1,φ_1 are the left and right eigenvectors of Eq. (3.59) corresponding to the eigenvalue μ_1. For convenience, we define

$$\chi(p) = \exp\left[-\int_{p_*}^{p} \mathcal{U}_1(y)\, dy\right] \tag{3.65}$$

and

$$\mathcal{B} = \mathbf{b}^T A^\dagger B\varphi_0 \tag{3.66}$$

where $A^\dagger$ is the pseudoinverse of A, φ_0 is the eigenvector spanning the nullspace of A, and $\mathbf{b}$ is as in Eq. (3.44). With all of these definitions in place, Newby & Keener [218] show that the escape eigenvalue λ_0 can be approximated as

$$\lambda_0 \approx \frac{\mathcal{B}(p_*)}{\pi\chi(p_\ell)} \sqrt{|\mathcal{U}_0''(p_\ell)|\mathcal{U}_0''(p_*)} e^{-N\mathcal{U}_0(p_\ell)} \tag{3.67}$$

For the systems discussed in Bressloff & Faugeras [216], Keener & Newby [217], Newby & Keener [218], each of the terms in Eq. (3.67) can be calculated analytically, either because of the simplicity of the model system or because of the low number of discrete states. Typical examples shown as validation of this theory have approximately 3 states to the Markov process. For our model system, the smallest number of possible states is 10^4. Regardless, each of the quantities can be estimated numerically to give a theoretical estimate for the escape time (Figure 16). We observe a reasonable fit between the theory estimate and the simulated escape times.

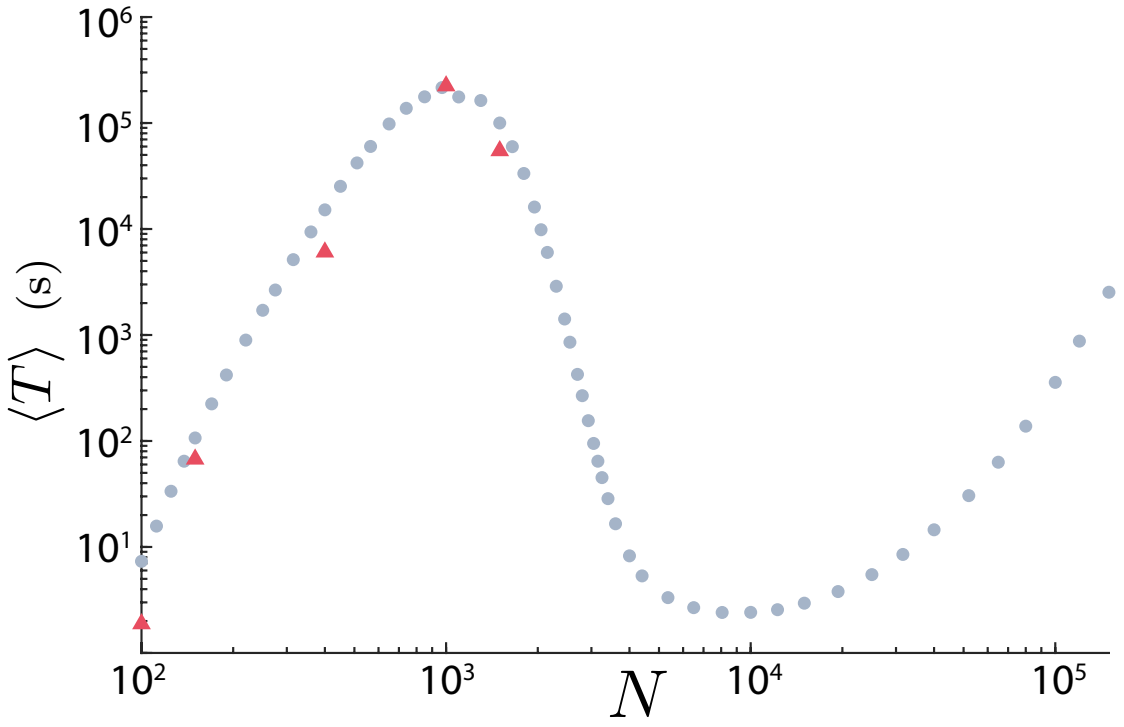

Figure 16: Theoretical estimate of escape time (red triangles) compared to simulated escape times (grey dots) of the model system Eq. (3.37). It is computationally infeasible at this time to estimate for $N > 1500$ (cf. section 3.5.3).

3.5.3 Eigenvalue estimation

To compute the estimates shown in section 3.5.2, we need to numerically estimate specific eigenvalues and eigenvectors of very large matrices. As this technique is independent of the biological modeling of the rest of this work, we discuss it in some generality.

We wish to solve the matrix equation

$$A\mathbf{y} = \mu B\mathbf{y}, \tag{3.68}$$

where we let $\mathbf{y}$ be the state space vector for the 2-dimensional Markov process, A the transition matrix for that process, and B the diagonal matrix of the $B_{(n,m)}(p)$ values denoting the right hand side of $\dot{p}_{IE}$ (Eq. (3.37), (3.38) and (3.41)). We recall that A,B are both functions of p. If we assume that B is invertible, as it will be for almost every value of p, Eq. (3.68) is equivalent to the standard eigenvalue problem

$$B^{-1}A\mathbf{y} = \mu\mathbf{y} \tag{3.69}$$

For notational simplicity, let us define $M_0 = B^{-1}A$. As discussed in section 3.5.2 and Newby & Keener [218], the matrix M_0 will have precisely two right eigenvectors that are also probability vectors. Our task is then to estimate these eigenvectors and the associated eigenvalues.

Generically, the process of computing eigenvalues and eigenvectors for an arbitrary $n \times n$ matrix M is well defined, and in fact analytically tractable. The eigenvalues are the roots of the characteristic polynomial, an n^{th} order polynomial in one variable. Though tedious, many simple methods for root-finding exist for one dimensional functions (simple gradient descent with a sufficient variety of initial conditions will suffice). Once the eigenvalues are obtained, the associated eigenvectors can be found by solving a matrix equation, through, for example, Gaussian row reduction. This process is ubiquitous throughout the applied sciences, and any undergraduate level textbook in linear algebra will contain descriptions of efficient algorithms for solving this problem (though of particular use are Golub & Van Loan [222] and Press et al. [223]). There are slightly more advanced techniques which are known to be efficient at computing eigenvalue/eigenvector pairs simultaneously, such as singular

value decomposition and QR-factorization[222, 224].

Though the process of eigenpair computation is well-defined, there are particular computational problems that arise when dealing with very large matrices. For our model system, with N neurons, recall that the two-dimensional Markov process has N^2 possible states, and so M_0 is of size $N^2 \times N^2$. Our numerical results span three orders of magnitude, with the smallest simulated system having $100^2 = 10^4$ possible states, and the largest simulated system having in excess of 10^{10} possible states. The symmetry properties of our model system ensure that the transition matrix A has a sparse representation of $\sim \mathcal{O}(N^2)$ nonzero entries, suggesting we may have some hope of solving this problem. However, the analytically motivated algorithms for computing eigenpairs cannot ensure sparseness of the factorization matrices. This means that, for example, in QR-factorization, the matrix Q will typically have $\sim \mathcal{O}(N^4)$ nonzero entries. For our model system, this means that a minimal Q requires storing 10^8 floating point numbers (which, at 4 bytes a float, accounts to $\sim$ 400MB to store the matrix), while the largest possible Q requires storing 10^{20} floating point numbers (which requires $\sim$ 100 million terabytes of storage). This is computationally infeasible.

To get around this, we can take advantage of the fact that we do not require the full eigenspectrum for our problem, but rather, only two specific eigenpairs. We make use of a technique know as *iterative subspace projection* (see Golub & Van Loan [222, Chapter 7]), which solves the eigenvalue problem for a *specific* eigenvalue by efficiently projecting the matrix onto subspaces. We note that a requirement for methods of this type are strong starting estimates for the eigenvalues. In practice, this is implemented using a canned function in MATLAB, `eigs` [225]. Even with the optimized algorithms of professional scientific software, memory constraints make computation infeasible for large systems. We are able to estimate the eigenpairs for up to 1500 neurons, but the combinatorial growth of required memory makes the current method of computation intractable for larger systems.

Then for us to effectively use this method, we need good estimates for our eigenvalues of interest. We know the eigenvalue $\mu_0 = 0$, and the associated eigenvector corresponds to the stationary distribution of the Markov process. For the eigenvalue μ_1 we do not have an *a priori* estimate. We do, however, know that $\mu_1(p)$ switches sign as p passes through the fixed points of Eq. (3.37). We can use this fact to construct an effective continuation method.

We adapt a method of eigenvalue continuation found in Kalaba et al. [226], though it seems this method is a variant of the Jacobi-Davidson algorithm discussed in Golub & Van Loan [222, (Chapter 10)], and is the basis for many continuation methods used in bifurcation theory[227].

Let M be an $n \times n$ matrix, and let $(\mathbf{y}, \mu)$ be an eigenpair, where M is parameterized by some parameter α. Furthermore, let $\mathbf{y} = \{y_i\}_{i=1}^n$ be a probability vector, i.e. $y_i \geq 0$ for $i \in \{1, \ldots, n\}$ and $\sum_i y_i = 1$. If we denote by $\mathbb{1}$ the $n \times 1$ vector of all ones, then this constrained eigenvalue problem may be stated as

$$\begin{aligned} M\mathbf{y} &= \mu\mathbf{y} \\ \mathbb{1}^T\mathbf{y} &= 1. \end{aligned} \tag{3.70}$$

Differentiating Eq. (3.70) with respect to α we find

$$\begin{aligned} \frac{\partial M}{\partial \alpha}\mathbf{y} + M\frac{\partial \mathbf{y}}{\partial \alpha} &= \frac{\partial \mu}{\partial \alpha}\mathbf{y} + \mu\frac{\partial \mathbf{y}}{\partial \alpha} \\ \mathbb{1}^T\frac{\partial \mathbf{y}}{\partial \alpha} &= 0 \end{aligned} \tag{3.71}$$

Re-arranging terms slightly, Eq. (3.71) becomes

$$\begin{aligned} \frac{\partial M}{\partial \alpha}\mathbf{y} &= (\mu I_n - M)\,\frac{\partial \mathbf{y}}{\partial \alpha} + \mathbf{y}\cdot\frac{\partial \mu}{\partial \alpha} \\ 0 &= \mathbb{1}^T\,\frac{\partial \mathbf{y}}{\partial \alpha} + 0\cdot\frac{\partial \mu}{\partial \alpha} \end{aligned} \tag{3.72}$$

where I_n is the $n \times n$ identity matrix. Letting differentiation with respect to α be denoted by $\prime$, Eq. (3.72) can be re-written as the matrix equation

$$\begin{pmatrix} M'\mathbf{y} \\ 0 \end{pmatrix} = \underbrace{\begin{pmatrix} \mu I_n - M & \mathbf{y} \\ \mathbb{1}^T & 0 \end{pmatrix}}_{:= J(p)} \begin{pmatrix} \mathbf{y}' \\ \mu' \end{pmatrix}, \tag{3.73}$$

which can otherwise be stated as the initial value problem

$$\begin{pmatrix} \mathbf{y}' \\ \mu' \end{pmatrix} = J^{-1}(p) \begin{pmatrix} M'\mathbf{y} \\ 0 \end{pmatrix}, \tag{3.74}$$

with initial conditions to be specified. As long as the matrix $J(p)$ is invertible, this problem

is well-defined, and we can use standard integration techniques to solve for the eigenpair $(\mathbf{y}, \mu)$.

To make sure that this problem is well defined, we should remark on the invertibility of $J(p)$. Rather than try to say something very general about $J(p)$, we instead look only at the restricted problem of interest to our system. We take $M = M_0 = B^{-1}A$, where A is the transition matrix defined by Eq. (3.38) and B is the diagonal matrix defined by Eq. (3.41). The Frobenius-Perron theorem guarantees that A has a unique eigenvector φ_0 with eigenvalue 0, with φ_0 being a probability vector (i.e. all positive terms, sums to 1). From Section 3.5.3, we know that M_0 has a unique eigenvector, φ_1, such that $\varphi_1 \neq \varphi_0$ and φ_1 is a probability vector. For values of p away from the fixed points of Eq. (3.37), $\mu_1(p) \neq 0$, and thus φ_1 and φ_0 are linearly independent of each other. Moreover, they will be linearly independent of any other eigenvectors (or generalized eigenvectors) of M_0. Since our goal is to estimate φ_0 and φ_1, it suffices to consider only a special case of J. Let J_1 be the matrix J with $\mathbf{y} = \varphi_1$, and similarly for J_0. Suppose $\mathbf{v}$ is a vector not in the span of φ_0,φ_1. Then $\mathbf{v}$ is expressible as a linear combination of eigenvectors and generalized eigenvectors of M_0, linearly independent of both φ_0 and φ_1. Then we observe that for any $a \in \mathbb{R}$, $\beta \in \{0, 1\}$,

$$J_\beta \begin{pmatrix} \mathbf{v} \\ a \end{pmatrix} = \begin{pmatrix} \mu_\beta \mathbf{v} - M_0 \mathbf{v} + a\varphi_\beta \\ \mathbb{1}^T \mathbf{v} \end{pmatrix}. \tag{3.75}$$

If the vector $(\mathbf{v}, a)^T$ is in the nullspace of J_β, we have

$$\mu_\beta \mathbf{v} - M_0 \mathbf{v} + a\varphi_\beta = 0, \tag{3.76}$$

which sets up a linear dependence between φ_β and the other eigenvectors and generalized eigenvectors of M_0. As both φ_0 and φ_1 span one dimensional eigenspaces, no such linear dependence can exist. Thus, J_β has a trivial kernel, thus is invertible, and the initial value problem in Eq. (3.71) is well-defined.

Because our problem is well-defined, solving Eq. (3.74) is in principle a trivial matter. Using a standard integration scheme, we compute the derivative term by solving Eq. (3.73) without explicitly inverting $J(p)$, and then adjust values and continue. This process can be entirely automated. In practice, however, the size of the matrices make this process

time-prohibitive, as solving a system of 10^{12} equations will tax even the best computers. For $N \leq 500$, we employ the automated process, solving Eq. (3.74) with a fixed-step fourth order Runge-Kutta integrator. For $N > 500$, we employ a partially automated process, where we estimate the derivative term in the integration scheme very crudely. Our algorithm is as follows:

1. Choose p near p_ℓ, so that $\mu_0 = 0$ and μ_1 near 0.
2. Using `eigs`, compute a set of the smallest magnitude eigenvalues of M_0 (typically ~ 10).
3. For each of the eigenvalues in the previous step, compute the eigenvectors. Precisely 2 will be probability vectors. These are our initial estimates for φ_0, φ_1, and μ_1.
4. Using `eigs`, refine the estimates from the previous step.
5. Compute the integration increment.
   a. If N is small (< 500), for a fixed increment Δp, compute the increment in μ_1,φ_1, φ_0 using Eq. (3.73).
   b. If N is large (> 500), for a fixed increment Δp, estimate the increment in μ_1 from the previous two values computed for μ_1. If we are on the n^{th} increment of p, then estimate the increment in μ_1 as the average of the slopes of the lines connecting $\mu_1(p_n)$ to $\mu_1(p_{n-1})$ and $\mu_1(p_{n-2})$.
6. Return to step 3 and iterate until p reaches a point past p_*.

By ranging over the interval of interest in the plasticity variable p, we are able to numerically approximate $\mu_1(p)$, as well as the left and right null vectors of $A - \mu_1 B$. Once we have this information, it is a simple matter to numerically estimate Eq. (3.63) to (3.66), as well as $\mathcal{U}_0''$. These are the estimates used in the generation of Figure 16.

3.6 CONCLUSION

Escape problems provide a useful perspective for studying a wide range of neuroscience phenomena, and were used here to understand state switching (between an asynchronous state and epochs of coordinated activity across the population) in the model of cortex first

presented in chapter 2. Following some initial introductory preliminaries on escape problems and bistable phenomena, we used a bistable Wilson-Cowan neuronal model as a canonical example to illustrate the basic ideas and approach, drawing out the similarities to the classic double-well potential.

We then examined our model of interest, a neuronal model with synaptic plasticity included. We compared the balanced network model used here with a traditionally scaled network, showing that the nonmonotonic relationship between escape time and system size in the balanced network is a characteristic not possible to replicate with traditional mean field scaling techniques. This nonmonotonicity appears to derive from the system size dependence in the governing equations, making analogous analysis to classic double well-type problems (i.e. using a potential energy function) more challenging.

To address this, we developed a psuedopotential, which allowed us to estimate escape times. This approach necessitated overcoming several significant computational hurdles, as the calculations needed to calculate escape times (without simulation) for even the smallest system ($N = 100$) involve 100^2 states. Nonetheless, for $N < 1500$, we were able to calculate escape times which closely matched simulation results for escape times using the same system.

In addition to the applications to cortex considered here, this work opens up several possible avenues for future research into the ways that finite size effects may explain population dynamics. These include both computational problems (such as efficient estimation of eigenvalues and eigenvectors for very large matrices), as well as analytical mathematical questions examining the proper way to study noise in systems where the potential landscape is intrinsically linked to the source of noise. In particular, much of the standard framework for studying dynamical behaviors in noisy systems is predicated on the explicit decoupling of the noise from an underlying dynamical landscape, making it possible to understand complex activity in terms of low dimensional systems. To effectively study systems where the dynamical landscape depends explicitly on the source of the noise, it will be necessary to extend our understanding of the current techniques of mean field models to include appropriate corrections.

4.0 CONCLUSION

In this work, we presented an extension of balanced neural network theory through the addition of short term synaptic plasticity and a treatment of stochastic effects. In chapter 2 we focused primarily on reproducing the qualitative features of an experimental result incongruous with simultaneous measurements. We were able to produce a balanced network that demonstrated rare events of large, population-wide correlated activity, generated by spontaneous stochastic effects. In chapter 3, we investigated the escape time behavior of a finite size balanced network, and observed strongly non-Arrhenius behavior in simulation. Following the work of Newby & Keener [218], we estimate the escape time by solving an appropriate boundary value problem, which necessitated the estimation of two eigenvectors of a matrix of $\mathcal{O}\left(N^4\right)$, with N between 10^2 and 10^6.

In the rest of this section, we detail a few of the several future directions and possible criticisms of this work. In both chapters, we focused on specific parameter ranges and values and a relatively simple model structure, to balance analytical tractability, parsimony, and realism. Thus, reasonable criticisms can be drawn with both the parameter values and particular choices for the model structure.

4.1 SPIKING SIMULATIONS

An essential continuation of this work is the construction of a spiking network model that reproduces the behavior we have outlined here. Qualitative models based on average quantities run the risk of either ignoring or introducing confounding factors that can lead to qualitative disagreement between the averaged model and the high dimensional spiking network. This

holds especially true for the work in chapter 3 as we are interpreting the transition inducing noise as due to finite size effects. This is the correct interpretation for the Markov model we present, but it is important to find validation with a spiking model that models the dynamics of N neurons, rather than modeling the dynamics of an aggregate of neurons (as we do in rate models).

There is a reasonable expectation that our results will be reproducible in spiking networks. For the system from chapter 2, the fact that the synapses are slow compared to both the excitatory and inhibitory populations means that the average synaptic activity must depend more on the firing rate than on specific spike times. This suggests that the qualitative features cannot be fine tuned, as there will be a large region of parameter space that produces statistically similar firing rate behavior. Moreover, balanced networks will have a broad distribution of firing rates[30, 54], and previous work has shown that population events of the type we study here are often a cascade effect instigated by ultra low frequency neurons [66] – essentially, neurons that rarely spike, but when they do, all the other neurons receive input. In that context, it seems likely that we will be able to produce events easily, and the slow rhythmic component can be independently tweaked by setting the timescales of the synapses. For the system from chapter 3, the change of synaptic scaling with system is essential to the novel result, and it is of interest to validate the non-monotonic nature of the escape time with the system size.

4.2 SYNAPTIC MODEL

An important question to pursue in future work is thus whether the behavior shown this work is robust to our choice of depression model. While the sigmoidal form used for the synaptic depression in chapter 2 simplifies and facilitates the bifurcation analysis, it differs from other models in the literature[57, 65, 66, 96]. More specifically, our choice of threshold steepness is particularly sharp. This large transition in synaptic efficacy over a small dynamic range in the rate variables serves to emphasize the switch-like behavior in our system. While this can be helpful from a modeling perspective, it disagrees somewhat with the known

biology[90, 100]. Recalling Figure 5B, we have access to direct measurements estimating the plasticity curve. To rectify this disagreement, we would wish to extend our model to include a more realistic model of synaptic plasticity.

To that end, as preliminary work, we have used a least squares fit to a Hill-type function to produce additional plasticity curves for use in the model (Figure 17 left panel). Using this plasticity curve fit to data, we are able to reproduce the basic features found in chapter 2, including persistent rhythmic activity, single events, and rhythmic events (Figure 17 right panels). There is a subtlety to this change of plasticity curve, since the rates in our model are scaled to live in $[0, 1]$ while the measured data from Oswald & Reyes [90], Oswald et al. [100] are in physical units. As such, we have an *effective threshold parameter* that controls the minimum rate at which depression is recruited while still retaining the least squares fits shown in Figure 17. Varying the effective threshold parameter is analogous to changing the thresholds θ_{EE},θ_{IE} in chapter 2, and is the main parameter we vary to produce the qualitatively different event types seen in Figure 17.

The next step is to implement a Tsodyks-Markram style synapse model, which is known to be able to faithfully reproduce synaptic behaviors observed in experiments. In particular, this work should be done in conjunction with the spiking network simulations discussed in section 4.1.

4.3 PARAMETER CHOICES

A broader exploration of the parameter values more generally would be valuable—for example, exploring the effects of different potential synaptic strengths on the model dynamics. A sensitivity analysis and exploration of all of the model parameters would allow us to evaluate whether our model can capture different behaviors, such as variations in the amplitude of population events. Related to the question of the potential range of behaviors in the model, in future work it may be useful to examine population event dynamics as a function of maturity (post-natal day) or cortical slice preparation (either by exploring the parameter space of this model, or by expanding the model to consider these issues).

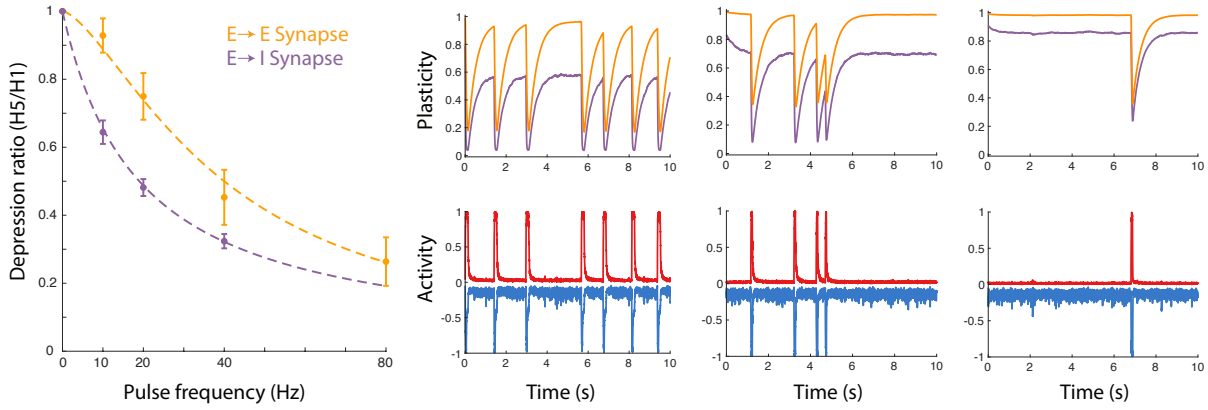

Figure 17: *Left panel:* Fit to data from [89, 90]. *Right panels:* Simulated event types from the model in chapter 2, with the plasticity variables replaced with the fit found in the left panel. Moving left to right, we see persistent oscillatory activity, rhythmic events, and singleton events. Along the top, plots of the two types of plasticity variables ($E \to I$ in purple, $E \to E$ in orange). Along the bottom, plots of the activity (excitatory activity in red, inhibitory activity in blue).

4.4 ESCAPE TIME ESTIMATE

Finally, we wish to improve on the escape time estimate. There are three main reasons why this is important. First, there are several standard techniques for computing escape times that we did not discuss in this work, and they were unsuccessful in reproducing the results found through simulation. A full understanding of why these methods failed is essential to fully understand the mechanism leading to the non-monotonicity. These other methods require computationally sensitive calculations, such as the estimating of a two dimensional unstable manifold, or the computation of a one-dimensional path in a six-dimensional space. These techniques are challenging at the best of times, and because of the nature of the large synaptic inputs in our model, these techniques become especially challenging for stiff systems. We have had some minor success at computing these paths and manifolds, but so far our solutions have not been robust enough to reliably reproduce the qualitative features of the escape time. Extending our work to successfully incorporate these additional estimates will add credence to our novel result.

Second, we note that the method we *have* successfully used to generate escape time estimates only works for systems sizes of up to ~ 1500. The computation of specific eigenvalues and eigenvectors of very large matrices is a difficult task, and one which has widespread applications in machine learning, as well as many areas of the applied sciences. Further advancement of eigenpair computations is generally of use, and adds to the skill set of applied researchers.

Third, while our method of escape time estimation provides good fits to the simulation data for the region we can calculate, the memory constraints on this method make it functionally impossible to numerically compute the escape times over the full range of simulated system sizes. Because the method we employ here effectively reduces our model to the escape of a single variable, we believe it should be possible to compute the escape times through a variety of different methods, and not just through the eigenpair estimation. There are a number of methods we have not had time to effectively implement, and we believe there are tractable solutions to this problem.

REFERENCES

[1] P. Dayan, L. F. Abbott & L. Abbott. Theoretical neuroscience: computational and mathematical modeling of neural systems (2001).

[2] G. B. Ermentrout & D. H. Terman. *Mathematical foundations of neuroscience*, volume 35 (Springer Science & Business Media, 2010).

[3] L. Abbott. Lapique's introduction of the integrate-and-fire model neuron (1907) brain res bull. 1999; 50: 303–304. doi: 10.1016. *S0361-9230 (99)* 00161–6 (1999).

[4] C. Morris & H. Lecar. Voltage oscillations in the barnacle giant muscle fiber. *Biophysical journal* **35**(1), 193–213 (1981).

[5] A. Shilnikov & M. Kolomiets. Methods of the qualitative theory for the hindmarsh–rose model: A case study–a tutorial. *International Journal of Bifurcation and chaos* **18**(08), 2141–2168 (2008).

[6] E. M. Izhikevich. Simple model of spiking neurons. *IEEE Transactions on neural networks* **14**(6), 1569–1572 (2003).

[7] R. Brette & W. Gerstner. Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of neurophysiology* **94**(5), 3637–3642 (2005).

[8] G. B. Ermentrout & N. Kopell. Parabolic bursting in an excitable system coupled with a slow oscillation. *SIAM Journal on Applied Mathematics* **46**(2), 233–253 (1986).

[9] A. L. Hodgkin & A. F. Huxley. The components of membrane conductance in the giant axon of loligo. *The Journal of physiology* **116**(4), 473–496 (1952).

[10] A. L. Hodgkin & A. F. Huxley. Currents carried by sodium and potassium ions through the membrane of the giant axon of loligo. *The Journal of physiology* **116**(4), 449–472 (1952).

[11] H. Ori, E. Marder & S. Marom. Cellular function given parametric variation in the hodgkin and huxley model of excitability. *Proceedings of the National Academy of Sciences* **115**(35), E8211–E8218 (2018).

[12] F. Khoyratee, T. Levi & S. Saïghi. Fpga implementation of the hodgkin-huxley model for neurological disease study. In *The 2nd International Symposium on Neuromorphic, non-linear, Neurofluidic Engineering, ISNNE* (2017).

[13] O. J. Walch & M. C. Eisenberg. Parameter identifiability and identifiable combinations in generalized hodgkin–huxley models. *Neurocomputing* **199**, 137–143 (2016).

[14] S. Rich, V. Booth & M. Zochowski. Intrinsic cellular properties and connectivity density determine variable clustering patterns in randomly connected inhibitory neural networks. *Frontiers in neural circuits* **10**, 82 (2016).

[15] N. Rashevsky. A suggestion for another statistical interpretation of the fundamental equations of the mathematical biophysics of the central nervous system. *The bulletin of mathematical biophysics* **7**(4), 223–226 (1945).

[16] W. R. Ashby. The stability of a randomly assembled nerve-network. *Electroencephalography and clinical neurophysiology* **2**(1-4), 471–482 (1950).

[17] W. R. Ashby, C. Walker et al. Instability of pulse activity in a net with threshold. *Nature* **196**, 561–562 (1962).

[18] R. L. Beurle. Properties of a mass of cells capable of regenerating pulses. *Phil. Trans. R. Soc. Lond. B* **240**(669), 55–94 (1956).

[19] J. Griffith. A field theory of neural nets: I: Derivation of field equations. *The bulletin of mathematical biophysics* **25**(1), 111–120 (1963).

[20] J. Griffith. On the stability of brain-like structures. *Biophysical journal* **3**(4), 299–308 (1963).

[21] H. R. Wilson & J. D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal* **12**(1), 1–24 (1972).

[22] G. Buzsáki. Large-scale recording of neuronal ensembles. *Nature neuroscience* **7**(5), 446 (2004).

[23] I. H. Stevenson & K. P. Kording. How advances in neural recording affect data analysis. *Nature neuroscience* **14**(2), 139 (2011).

[24] S.-i. Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological cybernetics* **27**(2), 77–87 (1977).

[25] G. B. Ermentrout & J. D. Cowan. A mathematical theory of visual hallucination patterns. *Biological cybernetics* **34**(3), 137–150 (1979).

[26] B. Ermentrout. Neural networks as spatio-temporal pattern-forming systems. *Reports on progress in physics* **61**(4), 353 (1998).

[27] A. Destexhe & T. J. Sejnowski. The wilson–cowan model, 36 years later. *Biological cybernetics* **101**(1), 1–2 (2009).

[28] D. J. Amit, H. Gutfreund & H. Sompolinsky. Spin-glass models of neural networks. *Physical Review A* **32**(2), 1007 (1985).

[29] N. Brunel. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of computational neuroscience* **8**(3), 183–208 (2000).

[30] A. Roxin, N. Brunel, D. Hansel, G. Mongillo & C. van Vreeswijk. On the distribution

of firing rates in networks of cortical neurons. *Journal of Neuroscience* **31**(45), 16217–16226 (2011).

[31] I. Ginzburg & H. Sompolinsky. Theory of correlations in stochastic neural networks. *Physical review E* **50**(4), 3171 (1994).

[32] S. H. Strogatz & R. E. Mirollo. Stability of incoherence in a population of coupled oscillators. *Journal of Statistical Physics* **63**(3-4), 613–635 (1991).

[33] R. E. Mirollo & S. H. Strogatz. Synchronization of pulse-coupled biological oscillators. *SIAM Journal on Applied Mathematics* **50**(6), 1645–1662 (1990).

[34] H. Sompolinsky, A. Crisanti & H.-J. Sommers. Chaos in random neural networks. *Physical review letters* **61**(3), 259 (1988).

[35] S. Coombes & Á. Byrne. Next generation neural mass models. In *Nonlinear Dynamics in Computational Neuroscience*, 1–16 (Springer, 2019).

[36] E. Ott & T. M. Antonsen. Low dimensional behavior of large systems of globally coupled oscillators. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **18**(3), 037113 (2008).

[37] G. Buzsáki & K. Mizuseki. The log-dynamic brain: how skewed distributions affect network operations. *Nature Reviews Neuroscience* **15**(4), 264–278 (2014).

[38] C. van Vreeswijk & H. Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* **274**(5293), 1724 (1996).

[39] R. J. Glauber. Time-dependent statistics of the ising model. *Journal of mathematical physics* **4**(2), 294–307 (1963).

[40] W. R. Softky & C. Koch. The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. *Journal of Neuroscience* **13**(1), 334–350 (1993).

[41] M. N. Shadlen & W. T. Newsome. Noise, neural codes and cortical organization. *Current opinion in neurobiology* **4**(4), 569–579 (1994).

[42] M. Graupner & A. D. Reyes. Synaptic input correlations leading to membrane potential decorrelation of spontaneous activity in cortex. *The Journal of Neuroscience* **33**(38), 15075–15085 (2013).

[43] Y. Shu, A. Hasenstaub & D. A. McCormick. Turning on and off recurrent balanced cortical activity. *Nature* **423**(6937), 288 (2003).

[44] B. Haider, A. Duque, A. R. Hasenstaub & D. A. McCormick. Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *Journal of Neuroscience* **26**(17), 4535–4545 (2006).

[45] M. Okun & I. Lampl. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature neuroscience* **11**(5), 535–537 (2008).

[46] M. R. DeWeese & A. M. Zador. Non-gaussian membrane potential dynamics imply sparse, synchronous activity in auditory cortex. *Journal of Neuroscience* **26**(47), 12206–12218 (2006).

[47] A. Luczak, P. Bartho & K. D. Harris. Gating of sensory input by spontaneous cortical activity. *Journal of Neuroscience* **33**(4), 1684–1695 (2013).

[48] J. Barral & A. D. Reyes. Synaptic scaling rule preserves excitatory-inhibitory balance and salient neuronal network dynamics. *Nature Neuroscience* **19**(12) (2016).

[49] L. T. Van Elst et al. Disturbed cingulate glutamate metabolism in adults with high-functioning autism spectrum disorder: evidence in support of the excitatory/inhibitory imbalance hypothesis. *Molecular psychiatry* **19**(12), 1314 (2014).

[50] J. R. Gibson, A. F. Bartley, S. A. Hays & K. M. Huber. Imbalance of neocortical excitation and inhibition and altered up states reflect network hyperexcitability in the

mouse model of fragile x syndrome. *Journal of neurophysiology* **100**(5), 2615–2626 (2008).

[51] R. Gao & P. Penzes. Common mechanisms of excitatory and inhibitory imbalance in schizophrenia and autism spectrum disorders. *Current molecular medicine* **15**(2), 146–167 (2015).

[52] M. L. Wallace, A. C. Burette, R. J. Weinberg & B. D. Philpot. Maternal loss of ube3a produces an excitatory/inhibitory imbalance through neuron type-specific synaptic defects. *Neuron* **74**(5), 793–800 (2012).

[53] F. Wendling, F. Bartolomei, J. Bellanger & P. Chauvel. Epileptic fast activity can be explained by a model of impaired gabaergic dendritic inhibition. *European Journal of Neuroscience* **15**(9), 1499–1508 (2002).

[54] A. Renart et al. The asynchronous state in cortical circuits. *science* **327**(5965), 587–590 (2010).

[55] C. van Vreeswijk & H. Sompolinsky. Chaotic balanced state in a model of cortical circuits. *Neural computation* **10**(6), 1321–1371 (1998).

[56] R. Rosenbaum & B. Doiron. Balanced networks of spiking neurons with spatially dependent recurrent connections. *Physical Review X* **4**(2), 021039 (2014).

[57] G. Mongillo, D. Hansel & C. van Vreeswijk. Bistability and spatiotemporal irregularity in neuronal networks with nonlinear synaptic transmission. *Physical review letters* **108**(15), 158101 (2012).

[58] A. D. Reyes. Synaptic short-term plasticity in auditory cortical circuits. *Hearing research* **279**(1), 60–66 (2011).

[59] R. S. Zucker & W. G. Regehr. Short-term synaptic plasticity. *Annual review of physiology* **64**(1), 355–405 (2002).

[60] H. Markram, Y. Wang & M. Tsodyks. Differential signaling via the same axon of neocortical pyramidal neurons. *Proceedings of the National Academy of Sciences* **95**(9), 5323–5328 (1998).

[61] B. Katz & R. Miledi. The role of calcium in neuromuscular facilitation. *The Journal of physiology* **195**(2), 481–492 (1968).

[62] R. Rosenbaum, J. Rubin & B. Doiron. Short term synaptic depression imposes a frequency dependent filter on synaptic information transfer. *PLoS computational biology* **8**(6), e1002557 (2012).

[63] J. De La Rocha & N. Parga. Short-term synaptic depression causes a non-monotonic response to correlated stimuli. *Journal of Neuroscience* **25**(37), 8416–8431 (2005).

[64] M. V. Tsodyks & H. Markram. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proceedings of the National Academy of Sciences* **94**(2), 719–723 (1997).

[65] L. F. Abbott, J. Varela, K. Sen & S. Nelson. Synaptic depression and cortical gain control. *Science* **275**(5297), 221–224 (1997).

[66] M. Tsodyks, A. Uziel, H. Markram et al. Synchrony generation in recurrent networks with frequency-dependent synapses. *J Neurosci* **20**(1), 825–835 (2000).

[67] M. Tsodyks, K. Pawelzik & H. Markram. Neural networks with dynamic synapses. *Neural computation* **10**(4), 821–835 (1998).

[68] T. P. Vogels, K. Rajan & L. F. Abbott. Neural network dynamics. *Annu. Rev. Neurosci.* **28**, 357–376 (2005).

[69] F. Wolf, R. Engelken, M. Puelma-Touzel, J. D. F. Weidinger & A. Neef. Dynamical models of cortical circuits. *Current opinion in neurobiology* **25**, 228–236 (2014).

[70] X.-J. Wang. Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological reviews* **90**(3), 1195–1268 (2010).

[71] C. Huang & B. Doiron. Once upon a (slow) time in the land of recurrent neuronal networks... *Current opinion in neurobiology* **46**, 31–38 (2017).

[72] S.-S. Poil, R. Hardstone, H. D. Mansvelder & K. Linkenkaer-Hansen. Critical-state dynamics of avalanches and oscillations jointly emerge from balanced excitation/inhibition in neuronal networks. *Journal of Neuroscience* **32**(29), 9817–9823 (2012).

[73] M. N. Shadlen & W. T. Newsome. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of neuroscience* **18**(10), 3870–3896 (1998).

[74] A. S. Ecker et al. Decorrelated neuronal firing in cortical microcircuits. *science* **327**(5965), 584–587 (2010).

[75] T. Herfurth & T. Tchumatchenko. How linear response shaped models of neural circuits and the quest for alternatives. *Current opinion in neurobiology* **46**, 234–240 (2017).

[76] B. Doiron, A. Litwin-Kumar, R. Rosenbaum, G. K. Ocker & K. Josić. The mechanics of state-dependent neural correlations. *Nature neuroscience* **19**(3), 383–393 (2016).

[77] T. Tetzlaff, M. Helias, G. T. Einevoll & M. Diesmann. Decorrelation of neural-network activity by inhibitory feedback. *PLoS computational biology* **8**(8), e1002596 (2012).

[78] M. Steriade, A. Nunez & F. Amzica. A novel slow (¡ 1 hz) oscillation of neocortical neurons in vivo: depolarizing and hyperpolarizing components. *Journal of neuroscience* **13**(8), 3252–3265 (1993).

[79] I. Timofeev, F. Grenier, M. Bazhenov, T. Sejnowski & M. Steriade. Origin of slow cortical oscillations in deafferented cortical slabs. *Cerebral Cortex* **10**(12), 1185–1199 (2000).

[80] J. M. Beggs & D. Plenz. Neuronal avalanches in neocortical circuits. *Journal of neuroscience* **23**(35), 11167–11177 (2003).

[81] T. Bellay, A. Klaus, S. Seshadri & D. Plenz. Irregular spiking of pyramidal neurons organizes as scale-invariant neuronal avalanches in the awake state. *Elife* **4**, e07224 (2015).

[82] A. Y. Tan, Y. Chen, B. Scholl, E. Seidemann & N. J. Priebe. Sensory stimulation shifts visual cortex from synchronous to asynchronous states. *Nature* **509**(7499), 226–229 (2014).

[83] G. Gigante, G. Deco, S. Marom & P. Del Giudice. Network events on multiple space and time scales in cultured neural networks and in a stochastic rate model. *PLoS Comput Biol* **11**(11), e1004547 (2015).

[84] A. Loebel & M. Tsodyks. Computation by ensemble synchronization in recurrent networks with synaptic depression. *Journal of computational neuroscience* **13**(2), 111–124 (2002).

[85] A. Loebel, I. Nelken & M. Tsodyks. Processing of sounds by population spikes in a model of primary auditory cortex. *Frontiers in neuroscience* **1**, 15 (2007).

[86] H. Ozeki, I. M. Finn, E. S. Schaffer, K. D. Miller & D. Ferster. Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron* **62**(4), 578–592 (2009).

[87] H. K. Kato, S. K. Asinof & J. S. Isaacson. Network-level control of frequency tuning in auditory cortex. *Neuron* **95**(2), 412–423 (2017).

[88] H. Adesnik. Synaptic mechanisms of feature coding in the visual cortex of awake mice. *Neuron* **95**(5), 1147–1159 (2017).

[89] A.-M. M. Oswald & A. D. Reyes. Maturation of intrinsic and synaptic properties of layer 2/3 pyramidal neurons in mouse auditory cortex. *Journal of neurophysiology* **99**(6), 2998–3008 (2008).

[90] A.-M. M. Oswald & A. D. Reyes. Development of inhibitory timescales in auditory cortex. *Cerebral Cortex* **21**(6), 1351–1361 (2011).

[91] G. Silberberg, C. Wu & H. Markram. Synaptic dynamics control the timing of neuronal excitation in the activated neocortical microcircuit. *The Journal of Physiology* **556**(1), 19–27 (2004).

[92] K. Vincent, J. S. Tauskela & J.-P. Thivierge. Extracting functionally feedforward networks from a population of spiking neurons. *Frontiers in computational neuroscience* **6** (2012).

[93] A. Luczak, B. L. McNaughton & K. D. Harris. Packet-based communication in the cortex. *Nature Reviews Neuroscience* **16**(12), 745–755 (2015).

[94] D. Millman, S. Mihalas, A. Kirkwood & E. Niebur. Self-organized criticality occurs in non-conservative neuronal networks during/up/'states. *Nature physics* **6**(10), 801–805 (2010).

[95] Z. P. Kilpatrick & P. C. Bressloff. Effects of synaptic depression and adaptation on spatiotemporal dynamics of an excitatory neuronal network. *Physica D: Nonlinear Phenomena* **239**(9), 547–560 (2010).

[96] E. Bart, S. Bao & D. Holcman. Modeling the spontaneous activity of the auditory cortex. *Journal of computational neuroscience* **19**(3), 357–378 (2005).

[97] M. V. Tsodyks, W. E. Skaggs, T. J. Sejnowski & B. L. McNaughton. Paradoxical effects of external modulation of inhibitory interneurons. *Journal of neuroscience* **17**(11), 4382–4388 (1997).

[98] R. B. Levy & A. D. Reyes. Spatial profile of excitatory and inhibitory synaptic connectivity in mouse primary auditory cortex. *Journal of Neuroscience* **32**(16), 5609–5619 (2012).

[99] T. Branco & K. Staras. The probability of neurotransmitter release: variability and

feedback control at single synapses. *Nature Reviews Neuroscience* **10**(5), 373–383 (2009).

[100] A.-M. M. Oswald, B. Doiron, J. Rinzel & A. D. Reyes. Spatial profile and differential recruitment of gabab modulate oscillatory activity in auditory cortex. *Journal of Neuroscience* **29**(33), 10321–10334 (2009).

[101] W. Horsthemke. Noise induced transitions. In *Non-Equilibrium Dynamics in Chemical Systems*, 150–160 (Springer, 1984).

[102] M. Mattia & P. Del Giudice. Finite-size dynamics of inhibitory and excitatory interacting spiking neurons. *Physical Review E* **70**(5), 052903 (2004).

[103] P. C. Bressloff. Metastable states and quasicycles in a stochastic wilson-cowan model of neuronal population dynamics. *Physical Review E* **82**(5), 051903 (2010).

[104] T. Schwalger, M. Deger & W. Gerstner. Towards a theory of cortical columns: From spiking neurons to interacting neural populations of finite size. *PLoS computational biology* **13**(4), e1005507 (2017).

[105] M. Benayoun, J. D. Cowan, W. van Drongelen & E. Wallace. Avalanches in a stochastic model of spiking neurons. *PLoS computational biology* **6**(7), e1000846 (2010).

[106] C. Curto, S. Sakata, S. Marguet, V. Itskov & K. D. Harris. A simple model of cortical dynamics explains variability and state dependence of sensory responses in urethane-anesthetized auditory cortex. *Journal of neuroscience* **29**(34), 10600–10612 (2009).

[107] S. Sakata & K. D. Harris. Laminar-dependent effects of cortical state on auditory cortical spontaneous activity. *Frontiers in neural circuits* **6**, 109 (2012).

[108] M. Pachitariu, D. R. Lyamzin, M. Sahani & N. A. Lesica. State-dependent population coding in primary auditory cortex. *Journal of Neuroscience* **35**(5), 2058–2073 (2015).

[109] M. Wehr & A. M. Zador. Synaptic mechanisms of forward suppression in rat auditory cortex. *Neuron* **47**(3), 437–445 (2005).

[110] M. Calford & M. Semple. Monaural inhibition in cat auditory cortex. *Journal of Neurophysiology* **73**(5), 1876–1891 (1995).

[111] M. Brosch & C. E. Schreiner. Time course of forward masking tuning curves in cat primary auditory cortex. *Journal of Neurophysiology* **77**(2), 923–943 (1997).

[112] M. Atzori et al. Differential synaptic processing separates stationary from transient inputs to the auditory cortex. *Nature neuroscience* **4**(12), 1230 (2001).

[113] G. B. Christianson, M. Sahani & J. F. Linden. Depth-dependent temporal response properties in core auditory cortex. *Journal of Neuroscience* **31**(36), 12837–12848 (2011).

[114] A. Reyes et al. Target-cell-specific facilitation and depression in neocortical circuits. *Nature neuroscience* **1**(4), 279–285 (1998).

[115] J. A. Varela et al. A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. *Journal of Neuroscience* **17**(20), 7926–7940 (1997).

[116] D. Hansel & C. van Vreeswijk. The mechanism of orientation selectivity in primary visual cortex without a functional map. *Journal of Neuroscience* **32**(12), 4049–4064 (2012).

[117] C. Pehlevan & H. Sompolinsky. Selectivity and sparseness in randomly connected balanced networks. *PLoS One* **9**(2), e89992 (2014).

[118] R. Rubin, L. Abbott & H. Sompolinsky. Balanced excitation and inhibition are required for high-capacity, noise-robust neuronal selectivity. *Proceedings of the National Academy of Sciences* 201705841 (2017).

[119] Y. Ahmadian, D. B. Rubin & K. D. Miller. Analysis of the stabilized supralinear network. *Neural computation* **25**(8), 1994–2037 (2013).

[120] O. Barak & M. Tsodyks. Working models of working memory. *Current opinion in neurobiology* **25**, 20–24 (2014).

[121] B. Haider, M. Häusser & M. Carandini. Inhibition dominates sensory responses in the awake cortex. *Nature* **493**(7430), 97–100 (2013).

[122] P. C. Bressloff. Stochastic neural field theory and the system-size expansion. *SIAM Journal on Applied Mathematics* **70**(5), 1488–1521 (2009).

[123] M. A. Buice & J. D. Cowan. Field-theoretic approach to fluctuation effects in neural networks. *Physical Review E* **75**(5), 051919 (2007).

[124] D. F. Anderson, B. Ermentrout & P. J. Thomas. Stochastic representations of ion channel kinetics and exact stochastic simulation of neuronal dynamics. *Journal of computational neuroscience* **38**(1), 67–82 (2015).

[125] B. Ermentrout. *Simulating, analyzing, and animating dynamical systems: a guide to XPPAUT for researchers and students*, volume 14 (Siam, 2002).

[126] K. D. Harris & A. Thiele. Cortical state and attention. *Nature reviews neuroscience* **12**(9), 509 (2011).

[127] I. Timofeev, F. Grenier & M. Steriade. Disfacilitation and active inhibition in the neocortex during the natural sleep-wake cycle: an intracellular study. *Proceedings of the National Academy of Sciences* **98**(4), 1924–1929 (2001).

[128] A. Luczak, P. Barthó, S. L. Marguet, G. Buzsáki & K. D. Harris. Sequential structure of neocortical spontaneous activity in vivo. *Proceedings of the National Academy of Sciences* **104**(1), 347–352 (2007).

[129] M. Steriade, A. Nuñez & F. Amzica. Intracellular analysis of relations between the slow (¡ 1 hz) neocortical oscillation and other sleep rhythms of the electroencephalogram. *Journal of Neuroscience* **13**(8), 3266–3283 (1993).

[130] D. S. Greenberg, A. R. Houweling & J. N. Kerr. Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nature neuroscience* **11**(7), 749 (2008).

[131] M. R. DeWeese & A. M. Zador. Non-gaussian membrane potential dynamics imply sparse, synchronous activity in auditory cortex. *Journal of Neuroscience* **26**(47), 12206–12218 (2006).

[132] S. Crochet & C. C. Petersen. Correlating whisker behavior with membrane potential in barrel cortex of awake mice. *Nature neuroscience* **9**(5), 608 (2006).

[133] J. F. Poulet & C. C. Petersen. Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. *Nature* **454**(7206), 881 (2008).

[134] A. Luczak, P. Barthó & K. D. Harris. Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron* **62**(3), 413–425 (2009).

[135] M. Okun, A. Naim & I. Lampl. The subthreshold relation between cortical local field potential and neuronal firing unveiled by intracellular recordings in awake rats. *Journal of neuroscience* **30**(12), 4440–4448 (2010).

[136] T. A. Engel et al. Selective modulation of cortical state during spatial attention. *Science* **354**(6316), 1140–1144 (2016).

[137] V. V. Vyazovskiy et al. Local sleep in awake rats. *Nature* **472**(7344), 443 (2011).

[138] R. Cossart, D. Aronov & R. Yuste. Attractor dynamics of network up states in the neocortex. *Nature* **423**(6937), 283 (2003).

[139] E. E. Fanselow & B. W. Connors. The roles of somatostatin-expressing (gin) and fast-spiking inhibitory interneurons in up-down states of mouse neocortex. *Journal of neurophysiology* **104**(2), 596–606 (2010).

[140] E. O. Mann, M. M. Kohl & O. Paulsen. Distinct roles of gabaa and gabab receptors in balancing and terminating persistent cortical activity. *Journal of Neuroscience* **29**(23), 7513–7518 (2009).

[141] M. V. Sanchez-Vives & D. A. McCormick. Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nature neuroscience* **3**(10), 1027 (2000).

[142] D. Plenz & S. T. Kitai. Up and down states in striatal medium spiny neurons simultaneously recorded with spontaneous activity in fast-spiking interneurons studied in cortex–striatum–substantia nigra organotypic cultures. *Journal of Neuroscience* **18**(1), 266–283 (1998).

[143] A. Compte et al. Spontaneous high-frequency (10–80 hz) oscillations during up states in the cerebral cortex in vitro. *Journal of Neuroscience* **28**(51), 13828–13844 (2008).

[144] K. T. Blackwell, U. Czubayko & D. Plenz. Quantitative estimate of synaptic inputs to striatal neurons during up and down states in vitro. *Journal of Neuroscience* **23**(27), 9123–9132 (2003).

[145] D. Contreras, I. Timofeev & M. Steriade. Mechanisms of long-lasting hyperpolarizations underlying slow sleep oscillations in cat corticothalamic networks. *The Journal of physiology* **494**(1), 251–264 (1996).

[146] S. Hill & G. Tononi. Modeling sleep and wakefulness in the thalamocortical system. *Journal of neurophysiology* **93**(3), 1671–1698 (2005).

[147] N. Parga & L. F. Abbott. Network model of spontaneous activity exhibiting synchronous transitions between up and down states. *Frontiers in Neuroscience* **1**, 4 (2007).

[148] M. Bazhenov, I. Timofeev, M. Steriade & T. J. Sejnowski. Model of thalamocortical slow-wave sleep oscillations and transitions to activated states. *Journal of neuroscience* **22**(19), 8691–8704 (2002).

[149] J. M. Benita, A. Guillamon, G. Deco & M. V. M. Sanchez-Vives. Synaptic depression and slow oscillatory activity in a biophysical network model of the cerebral cortex. *Frontiers in computational neuroscience* **6**, 64 (2012).

[150] A. Compte et al. Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *Journal of neurophysiology* **90**(5), 3441–3454 (2003).

[151] J.-Y. Chen, S. Chauvette, S. Skorheim, I. Timofeev & M. Bazhenov. Interneuron-mediated inhibition synchronizes neuronal activity during slow oscillation. *The Journal of physiology* **590**(16), 3987–4010 (2012).

[152] I. A. Erchova, M. A. Lebedev & M. E. Diamond. Somatosensory cortical neuronal population activity across states of anaesthesia. *European Journal of Neuroscience* **15**(4), 744–752 (2002).

[153] S. Chauvette, M. Volgushev & I. Timofeev. Origin of active states in local neocortical networks during slow sleep oscillation. *Cerebral cortex* **20**(11), 2660–2674 (2010).

[154] I. Lampl, I. Reichova & D. Ferster. Synchronous membrane potential fluctuations in neurons of the cat visual cortex. *Neuron* **22**(2), 361–374 (1999).

[155] R. Beltramo et al. Layer-specific excitatory circuits differentially control recurrent network dynamics in the neocortex. *Nature neuroscience* **16**(2), 227 (2013).

[156] A. Stroh et al. Making waves: initiation and propagation of corticothalamic ca2+ waves in vivo. *Neuron* **77**(6), 1136–1150 (2013).

[157] J. Tabak, J. Rinzel & R. Bertram. Quantifying the relative contributions of divisive and subtractive feedback to rhythm generation. *PLoS computational biology* **7**(4), e1001124 (2011).

[158] J. Tabak, W. Senn, M. J. O'Donovan & J. Rinzel. Modeling of spontaneous activity in developing spinal cord using activity-dependent depression in an excitatory network. *Journal of Neuroscience* **20**(8), 3041–3056 (2000).

[159] D. Holcman & M. Tsodyks. The emergence of up and down states in cortical networks. *PLoS computational biology* **2**(3), e23 (2006).

[160] A. Destexhe. Self-sustained asynchronous irregular states and up–down states in thalamic, cortical and thalamocortical networks of nonlinear integrate-and-fire neurons. *Journal of computational neuroscience* **27**(3), 493 (2009).

[161] D. Jercog et al. Up-down cortical dynamics reflect state transitions in a bistable network. *Elife* **6**, e22425 (2017).

[162] L. Parkkonen, J. Andersson, M. Hämäläinen & R. Hari. Early visual brain areas reflect the percept of an ambiguous scene. *Proceedings of the National Academy of Sciences* **105**(51), 20500–20504 (2008).

[163] J. Brascamp, P. Klink & W. J. Levelt. The 'laws' of binocular rivalry: 50 years of levelt's propositions. *Vision research* **109**, 20–37 (2015).

[164] W. J. Levelt. *On binocular rivalry*. Ph.D. thesis, Van Gorcum Assen (1965).

[165] M. Tervaniemi, S. Maury & R. Näätänen. Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *Neuroreport* **5**(7), 844–846 (1994).

[166] C. Murphy, W. S. Cain & L. M. Bartoshuk. Mutual action of taste and olfaction. *Sensory processes* (1977).

[167] J. A. Gottfried & R. J. Dolan. The nose smells what the eye sees: crossmodal visual facilitation of human olfactory perception. *Neuron* **39**(2), 375–386 (2003).

[168] W. Hill. My wife and my mother-in-law. *Puck* **16**, 11 (1915).

[169] I. C. McManus, M. Freegard, J. Moore & R. Rawles. Science in the making: Right hand, left hand. ii: The duck–rabbit figure. *Laterality* **15**(1-2), 166–185 (2010).

[170] L. C. Lack. Selective attention and the control of binocular rivalry. *Perception & Psychophysics* **15**(1), 193–200 (1974).

[171] A. Kleinschmidt, C. Büchel, S. Zeki & R. S. Frackowiak. Human brain activity during spontaneously reversing perception of ambiguous figures. *Proceedings of the Royal Society of London B: Biological Sciences* **265**(1413), 2427–2433 (1998).

[172] D. A. Leopold & N. K. Logothetis. Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* **379**(6565), 549 (1996).

[173] A. J. Parker & K. Krug. Neuronal mechanisms for the perception of ambiguous stimuli. *Current opinion in neurobiology* **13**(4), 433–439 (2003).

[174] S. C. Chong, D. Tadin & R. Blake. Endogenous attention prolongs dominance durations in binocular rivalry. *Journal of vision* **5**(11), 6–6 (2005).

[175] G. J. Brouwer & R. van Ee. Endogenous influences on perceptual bistability depend on exogenous stimulus characteristics. *Vision Research* **46**(20), 3393 – 3402 (2006).

[176] K. Hol, A. Koene & R. van Ee. Attention-biased multi-stable surface perception in three-dimensional structure-from-motion. *Journal of vision* **3**(7), 3–3 (2003).

[177] T. C. Toppino. Reversible-figure perception: Mechanisms of intentional control. *Perception & psychophysics* **65**(8), 1285–1295 (2003).

[178] L. C. van Dam & R. van Ee. The role of saccades in exerting voluntary control in perceptual and binocular rivalry. *Vision research* **46**(6-7), 787–799 (2006).

[179] R. van Ee, L. Van Dam & G. Brouwer. Voluntary control and the dynamics of perceptual bi-stability. *Vision research* **45**(1), 41–55 (2005).

[180] J. W. Brascamp & R. Blake. Inattention abolishes binocular rivalry: Perceptual evidence. *Psychological Science* **23**(10), 1159–1167 (2012).

[181] J. W. Brascamp, R. Van Ee, W. R. Pestman & A. V. Van Den Berg. Distributions of alternation rates in various forms of bistable perception. *Journal of Vision* **5**(4), 1–1 (2005).

[182] R. M. Bote, J. Rinzel & N. Rubin. Noise-induced alternations in an attractor network model of perceptual bi-stability. *Journal of neurophysiology* (2017).

[183] A. Shpiro, R. Moreno-Bote, N. Rubin & J. Rinzel. Balance between noise and adaptation in competition models of perceptual bistability. *Journal of computational neuroscience* **27**(1), 37 (2009).

[184] M.-S. Kang & R. Blake. What causes alternations in dominance during binocular rivalry? *Attention, Perception, & Psychophysics* **72**(1), 179–186 (2010).

[185] R. Moreno-Bote, J. Rinzel & N. Rubin. Noise-induced alternations in an attractor network model of perceptual bistability. *Journal of neurophysiology* **98**(3), 1125–1139 (2007).

[186] A. Shpiro, R. Curtu, J. Rinzel & N. Rubin. Dynamical characteristics common to neuronal competition models. *Journal of neurophysiology* **97**(1), 462–473 (2007).

[187] L. Stollenwerk & M. Bode. Lateral neural model of binocular rivalry. *Neural computation* **15**(12), 2863–2882 (2003).

[188] C. Huang, B. Englitz, S. Shamma & J. Rinzel. A neuronal network model for context-dependence of pitch change perception. *Frontiers in computational neuroscience* **9**, 101 (2015).

[189] W. Maass. On the computational power of winner-take-all. *Neural computation* **12**(11), 2519–2535 (2000).

[190] R. Coultrip, R. Granger & G. Lynch. A cortical model of winner-take-all competition via lateral inhibition. *Neural networks* **5**(1), 47–54 (1992).

[191] S. Kaski & T. Kohonen. Winner-take-all networks for physiological models of competitive learning. *Neural Networks* **7**(6-7), 973–984 (1994).

[192] J. M. Beggs & D. Plenz. Neuronal avalanches in neocortical circuits. *Journal of neuroscience* **23**(35), 11167–11177 (2003).

[193] J. M. Palva et al. Neuronal long-range temporal correlations and avalanche dynamics are correlated with behavioral scaling laws. *Proceedings of the National Academy of Sciences* **110**(9), 3585–3590 (2013).

[194] J. M. Beggs & D. Plenz. Neuronal avalanches are diverse and precise activity patterns that are stable for many hours in cortical slice cultures. *Journal of neuroscience* **24**(22), 5216–5229 (2004).

[195] T. Petermann et al. Spontaneous cortical activity in awake monkeys composed of neuronal avalanches. *Proceedings of the National Academy of Sciences* **106**(37), 15921–15926 (2009).

[196] D. Plenz & T. C. Thiagarajan. The organizing principles of neuronal avalanches: cell assemblies in the cortex? *Trends in neurosciences* **30**(3), 101–110 (2007).

[197] V. Pasquale, P. Massobrio, L. Bologna, M. Chiappalone & S. Martinoia. Self-organization and neuronal avalanches in networks of dissociated cortical neurons. *Neuroscience* **153**(4), 1354–1369 (2008).

[198] W. L. Shew, H. Yang, T. Petermann, R. Roy & D. Plenz. Neuronal avalanches imply maximum dynamic range in cortical networks at criticality. *Journal of neuroscience* **29**(49), 15595–15600 (2009).

[199] W. L. Shew, H. Yang, S. Yu, R. Roy & D. Plenz. Information capacity and transmission are maximized in balanced cortical networks with neuronal avalanches. *Journal of neuroscience* **31**(1), 55–63 (2011).

[200] R. M. Mazo. *Brownian motion: fluctuations, dynamics, and applications*, volume 112 (Oxford University Press on Demand, 2002).

[201] A. Einstein. On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat. *Annalen der physik* **17**, 549–560 (1905).

[202] N. Wiener. The average of an analytic functional and the brownian movement. *Proceedings of the National Academy of Sciences* **7**(10), 294–298 (1921).

[203] N. Wiener. *Cybernetics or Control and Communication in the Animal and the Machine*, volume 25 (MIT press, 1961).

[204] N. Wiener. *The human use of human beings: Cybernetics and society*. 320 (Perseus Books Group, 1988).

[205] H. Risken. Fokker-planck equation. In *The Fokker-Planck Equation*, 63–95 (Springer, 1996).

[206] C. Gardiner. *Stochastic methods*, volume 4 (springer Berlin, 2009).

[207] M. C. Wang & G. E. Uhlenbeck. On the theory of the brownian motion ii. *Reviews of modern physics* **17**(2-3), 323 (1945).

[208] B. J. Matkowsky & Z. Schuss. The exit problem for randomly perturbed dynamical systems. *SIAM Journal on Applied Mathematics* **33**(2), 365–382 (1977).

[209] Z. Schuss & B. J. Matkowsky. The exit problem: a new approach to diffusion across potential barriers. *SIAM Journal on Applied Mathematics* **36**(3), 604–623 (1979).

[210] B. Matkowsky, Z. Schuss & C. Tier. Uniform expansion of the transition rate in kramers' problem. *Journal of statistical physics* **35**(3-4), 443–456 (1984).

[211] P. Hänggi, P. Talkner & M. Borkovec. Reaction-rate theory: fifty years after kramers. *Reviews of modern physics* **62**(2), 251 (1990).

[212] B. Matkowsky & Z. Schuss. Eigenvalues of the fokker–planck operator and the approach to equilibrium for diffusions in potential fields. *SIAM Journal on Applied Mathematics* **40**(2), 242–254 (1981).

[213] B. Matkowsky & Z. Schuss. Diffusion across characteristic boundaries. *SIAM Journal on Applied Mathematics* **42**(4), 822–834 (1982).

[214] B. J. Matkowsky, A. Nitzan & Z. Schuss. Does reaction path curvature play a role in the diffusion theory of multidimensional activated rate processes? *The Journal of chemical physics* **88**(8), 4765–4771 (1988).

[215] E. Ben-Jacob, D. Bergman, B. Matkowsky & Z. Schuss. Lifetime of oscillatory steady states. *Physical Review A* **26**(5), 2805 (1982).

[216] P. C. Bressloff & O. Faugeras. On the hamiltonian structure of large deviations in stochastic hybrid systems. *Journal of Statistical Mechanics: Theory and Experiment* **2017**(3), 033206 (2017).

[217] J. P. Keener & J. M. Newby. Perturbation analysis of spontaneous action potential initiation by stochastic ion channels. *Physical Review E* **84**(1), 011918 (2011).

[218] J. M. Newby & J. P. Keener. An asymptotic analysis of the spatially inhomogeneous velocity-jump process. *Multiscale Modeling & Simulation* **9**(2), 735–765 (2011).

[219] P. C. Bressloff & S. D. Lawley. Mean first passage times for piecewise deterministic markov processes and the effects of critical points. *Journal of Statistical Mechanics: Theory and Experiment* **2017**(6), 063202 (2017).

[220] J.-Y. Lee & M. J. Ward. On the asymptotic and numerical analyses of exponentially iii-conditioned singularly perturbed boundary value problems. *Studies in Applied Mathematics* **94**(3), 271–326 (1995).

[221] C. M. Bender & S. A. Orszag. *Advanced mathematical methods for scientists and engineers I: Asymptotic methods and perturbation theory* (Springer Science & Business Media, 2013).

[222] G. H. Golub & C. F. Van Loan. *Matrix computations*, volume 3 (JHU Press, 2012).

[223] W. H. Press, S. A. Teukolsky, W. T. Vetterling & B. P. Flannery. *Numerical recipes 3rd edition: The art of scientific computing* (Cambridge university press, 2007).

[224] R. A. Horn & C. R. Johnson. *Matrix analysis* (Cambridge university press, 1990).

[225] MATLAB. *version 9.4.0 (R2018a)* (The MathWorks Inc., Natick, Massachusetts, 2018).

[226] R. Kalaba, K. Spingarn & L. Tesfatsion. Individual tracking of an eigenvalue and eigenvector of a parameterized matrix. *Nonlinear Analysis: Theory, Methods & Applications* **5**(4), 337–340 (1981).

[227] Y. A. Kuznetsov. *Elements of applied bifurcation theory*, volume 112 (Springer Science & Business Media, 2013).